

DOCUMENT RESUME

ED 041 270

AL 002 472

AUTHOR Heenan, David K.
TITLE Evaluation of Instruction in Thailand. Education in Thailand Publication Number 1.
INSTITUTION Michigan State Univ., East Lansing.; Ministry of Education, Bangkok (Thailand).
SPONS AGENCY Agency for International Development, Washington, D.C.
PUB DATE 66
NOTE 117p.
AVAILABLE FROM Institute for International Studies in Education, College of Education, Michigan State University, East Lansing, Michigan 48823

EDRS PRICE MF-\$0.50 HC-\$5.95
DESCRIPTORS Developing Nations, *Evaluation, Evaluation Methods, *Evaluation Techniques, Guidelines, *Measurement Instruments, Scoring, Student Evaluation, *Teacher Education, *Test Construction, Testing, Thai
IDENTIFIERS *Thailand

ABSTRACT

In 1964 the National Economic Development Board of the Government of Thailand and the U.S. Agency for International Development entered into an agreement with Michigan State University to provide technical assistance to certain educational agencies in Thailand in the area of educational planning. A program of professional publications was completed entitled the "Education in Thailand Series." The primary objectives of the series are: (1) to provide a means whereby interested individuals may obtain a digest of major cooperative research projects conducted by Thai and U.S. personnel, (2) to prepare material which will introduce selected educational concepts to personnel of the Thai educational system, (3) to review and translate the findings of other pertinent research projects conducted by the participating staff of Michigan State while in Thailand, and (4) to create a framework for educational publications which will interpret developments in Thai education to the educators of the world. This publication was the first in the series and was compiled to provide teachers in Thailand with information about evaluation and evaluative instruments which might be useful in their classroom work. It is written in Thai and English and includes: Evaluation and Evaluative Instruments, Types of Measuring Instruments, Constructing Examinations, Glossary. (JD)

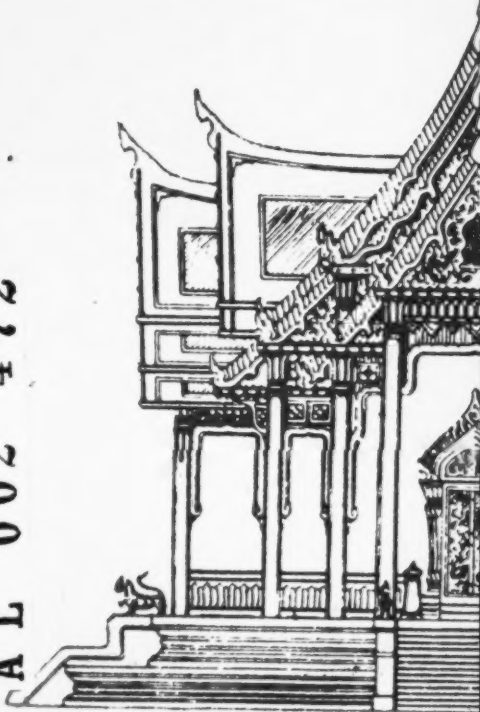
ED 004 1270

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

EDUCATION IN THAILAND

AL 002 472



EVALUATION OF INSTRUCTION IN THAILAND

B/c

การศึกษา
ในประเทศไทย

การประเมินผล
การสอน
ในประเทศไทย

ED041270

การประเมินผลการสอนในประเทศไทย

เอกสารอันดับที่ ๑

ศาสตราจารย์ เดวิด เค. ฮีแนน

มหาวิทยาลัยมิชิแกนสเตต

บรรณาธิการฝ่ายไทย ดร. ชวาล แพร่ตกุล

ผู้อำนวยการสำนักงานทดสอบ

วิทยาลัยวิชาการศึกษาประสานมิตร

EVALUATION OF INSTRUCTION IN THAILAND

Publication Number 1

David K. Heenan

PROFESSOR

MICHIGAN STATE UNIVERSITY

Thai editor: Chaval Paeratakul

DIRECTOR

TEST BUREAU

COLLEGE OF EDUCATION

PRASARNMITR

AL 002 472

การศึกษาในประเทศไทย

ในปี ค.ศ. 1964 สภาพัฒนาการเศรษฐกิจแห่งชาติ (NEDB) และองค์การเพื่อการพัฒนาระหว่างประเทศ (AID) แห่งสหรัฐอเมริกาได้ร่วมกันตกลงให้ Michigan State University เป็นผู้จัดการให้ความช่วยเหลือทางด้านวิชาการแก่หน่วยราชการที่เกี่ยวข้องกับการศึกษาต่าง ๆ ในประเทศไทย การให้ความช่วยเหลือนี้ ได้กำหนดแนวนอนลงไว้ในเรื่องการวางแผนการศึกษา และในการดำเนินการทั้งหลาย สำนักงานวางแผนการศึกษาของกระทรวงศึกษาธิการ และสำนักงานสภาการศึกษาแห่งชาติ ได้ตกลงที่จะให้ความร่วมมืออย่างใกล้ชิด ในข้อตกลงอันนี้ มีโครงการที่จะจัดพิมพ์เอกสารชุด ซึ่งมีชื่อเรียก "การศึกษาในประเทศไทย"

ความมุ่งหมายเริ่มแรกในการจัดพิมพ์เอกสารชุดนี้ คือ

1. เพื่อเป็นสื่อกลางให้ผู้สนใจใคร่ทราบโดยสังเขป ถึงโครงการเกี่ยวกับการวางแผนการศึกษาที่สำคัญ ๆ ซึ่งเจ้าหน้าที่ฝ่ายไทยและอเมริกัน กำลังร่วมกันดำเนินงานอยู่
2. เพื่อประมวลเป็นเอกสารที่จะรวมแนวความคิดเกี่ยวกับการศึกษาโดยทั่ว ๆ ไปให้แก่เจ้าหน้าที่บริหารการศึกษาฝ่ายไทย
3. พิจารณาและวิจารณ์ข้อมูลสำคัญต่าง ๆ ที่คณะ Michigan State University ได้พบในการที่มีส่วนร่วมในโครงการต่าง ๆ เกี่ยวกับการวางแผนการศึกษา ขณะที่ประจำอยู่ในประเทศไทย
4. วางแนวทางในการจัดพิมพ์เอกสารทางด้านการศึกษา ซึ่งจะแสดงพัฒนาการทางการศึกษาของประเทศไทยให้เป็นที่เข้าใจแก่นักบริหารการศึกษาในวงการศึกษาทั่วโลก

เอกสารนี้เป็นเรื่องหนึ่งในเอกสารชุดที่จะพิมพ์ เอกสารนี้ได้เรียบเรียงขึ้นตามความมุ่งหมายที่กล่าวแล้วข้างต้น และหวังว่าข้อเขียนนี้จะมีส่วนช่วยบ้างในการส่งเสริมการปรับปรุงระบบการศึกษาของประเทศไทย และช่วยให้ประเทศอื่น ๆ มีความรู้และความเข้าใจพัฒนาการทางการศึกษาของประเทศไทยดีขึ้น

EDUCATION IN THAILAND

In 1964 the National Economic Development Board of the Government of Thailand and the U.S. Agency for International Development entered into an agreement with Michigan State University to provide technical assistance to certain educational agencies in Thailand. The general purpose for the assistance was determined to be in the area of educational planning and it was decided that it should be done through the Educational Planning Office of the Ministry of Education and the National Education Council. A program of professional publications was made possible by this agreement and has been entitled the **Education in Thailand** series.

The primary objectives of the series are:

1. To provide a means whereby interested individuals may obtain a digest of major cooperative research projects conducted by Thai and U.S. personnel.
2. To prepare material which will introduce selected educational concepts to personnel of the Thai educational system.
3. To review and translate the findings of other pertinent research projects conducted by the participating staff of Michigan State University while in Thailand.
4. To create a framework for educational publications which will interpret developments in Thai education to the educators of the world.

This publication is one of the titles in the series. It has been prepared in accordance with the objectives noted above and has been written in the hope that in a small way it will contribute to a better educational system for Thailand and assist other countries to gain an understanding of educational developments in the Kingdom.

คำนำ

เอกสารฉบับนี้ได้จัดทำขึ้นเพื่อรวบรวมข้อมูลเกี่ยวกับการประเมินผล และเครื่องมือประเมินผลที่ใช้ในชั้นเรียน เพื่อเป็นประโยชน์แก่ผู้ทำการสอนทั้งหลายในประเทศไทย ในการจัดพิมพ์เอกสารนี้ ได้มีผู้กรุณาให้ความร่วมมือและช่วยเหลือหลายท่านด้วยกัน ผู้ที่ช่วยเหลือในการให้ข้อคิดเห็นและการจัดพิมพ์เป็นภาษาอังกฤษได้แก่ ผู้เชี่ยวชาญทางการศึกษาตามโครงการ ช่วยเหลือทางการวางแผนการศึกษาในประเทศไทย ของมหาวิทยาลัย Michigan State University สองท่าน คือ Dr. Raymond N. Hatch หัวหน้าคณะ และ Dr. Stanley P. Wronski ที่ปรึกษาทางการศึกษาของสำนักงานวางแผนการศึกษา กระทรวงศึกษาธิการ นอกจากนี้ ดร. ชวาล แพรัตกุล หัวหน้าสำนักงานทดสอบ ของวิทยาลัยวิชาการศึกษา ประสานมิตร และผู้ช่วยอีกสองท่าน คือ นายวัน สังข์สะอาดและ นายสวัสดิ์ ประทุมราช ได้กรุณาให้ความช่วยเหลือในการจัดแปลเป็นภาษาไทย และให้ข้อคำถาม (Test items) อื่นๆ อีกมากมาย อีกประการหนึ่ง น.ส. สมบูรณ์ บุณยวิทย์ แห่งสำนักงานสภาการศึกษาแห่งชาติ ได้กรุณาเขียนภาพประกอบต่าง ๆ ในเอกสารฉบับนี้ให้ ผู้เขียนยังถือโอกาสขอบคุณเจ้าหน้าที่สำนักงาน Evaluation Services แห่ง Michigan State University ที่ได้ส่งเอกสารและความคิดเห็นมาประกอบการจัดพิมพ์เอกสารนี้ด้วย

David K. Heenan

กรุงเทพ ฯ ประเทศไทย

2509

PREFACE

This brochure was compiled to provide teachers in Thailand with information about evaluation and evaluative instruments which might be useful in their classroom work. Many people have contributed suggestions and assistance in the preparation of this manuscript: Dr. Raymond N. Hatch, Chief of Party, and Dr. Stanley P. Wronski, Educational Consultant to the Educational Planning Office of the Ministry of Education, both members of the Michigan State Contract Team in Thailand, have helped with the editing of the English part of the text. Dr. Chaval Paeratakul, Director of the Test Bureau at the College of Education, Prasarnmitr, and his assistants, Mr. Wan Sangsa-ard and Mr. Sawat Pratoomraj, provided the Thai translation and many of the test items. The illustrations are the work of Mrs. Suwimol Buah-Apukdee, from the National Education Council. The writer also wishes to acknowledge the materials and ideas obtained from his colleagues in the Office of Evaluation Services at Michigan State University.

David K. Heenan
Bangkok, Thailand
1966

สารบัญ

	หน้า
การประเมินค่าและเครื่องมือของการประเมินค่า	1
การประเมินค่าและความมุ่งหมาย	2
การประเมินค่าและการจูงใจนักเรียน	3
การสอบและการให้เกรด	3
การประเมินค่าเป็นเครื่องมือของการเรียน	4
การประเมินค่าและการเข้าใจตัวเด็ก	5
สรุปความ	5
II ประเภทของเครื่องมือวัดผล	6
การสอบแบบปากเปล่า	6
การสอบแบบข้อเขียน	10
III การสร้างข้อสอบ	15
การสอบปากเปล่า	17
การสอบข้อเขียนแบบเรียงความ	19
ข้อสอบแบบปรนัย	27
การดำเนินการทดสอบ	48
การตรวจและการวิเคราะห์ผลสอบ	50
อภิธานศัพท์	55

Table of Contents

	Page
I Evaluation and Evaluative Instruments	59
Evaluation and Objectives	59
Evaluation and Student Motivation	60
Examinations and Grades	60
Evaluation as a Learning Instrument	61
Evaluation and Understanding Students	61
 II Types of Measuring Instruments	62
Oral Tests	63
Written Examinations	65
 III Constructing Examinations	69
Oral Examinations	70
Written Examinations: The Essay	72
Objective Tests	79
Administering the Examination	97
Scoring and Analyzing the Examination	98
Glossary of Terms	103

I การประเมินค่า และเครื่องมือของการประเมินค่า

ชาวไทยมองเห็นวัตถุประสงค์ และความสำคัญของการประเมินค่า ตามความหมายใหม่ เมื่อไม่กี่ปีมานี้เอง ในสมัยที่ระบบการศึกษายังอยู่ใน วงแคบ และประเทศชาติก็ยังมุ่งมั่นต่อการพัฒนาน้อยอยู่นั้น การนำเอา ขบวนการศึกษาที่ต้องสิ้นเปลืองสูง (และมักจะได้ผลน้อย) มาใช้ ก็ยังอาจ จะกระทำกันอยู่ได้ แต่ในปัจจุบัน ขณะที่รัฐกำลังเน้นหนักต่อการพัฒนา ประเทศชาติตามเป้าหมายใหม่นี้ ได้ก่อให้เกิดความจำเป็นและความต้องการ ใหม่ๆ เพิ่มขึ้นอีกนานัปการ ทั้งเช่นการขยายการศึกษาภาคบังคับให้นาน ออกไป และการเพิ่มปริมาณอย่างมากมายของเด็กที่จะต้องเข้าเรียนเป็นต้น สิ่งเหล่านี้ทำให้ครู นักบริหาร และผู้นำของรัฐ ต้องหันมาพิจารณา กิจกรรมที่กำลังดำเนินอยู่ในขณะนี้ อย่างจริงจังว่า ระบบการศึกษาใน ปัจจุบัน เป็นขบวนการที่ประหยัดและมีประสิทธิภาพสมกับเงินงบประมาณ ที่ได้รับอยู่เพียงใด การวิเคราะห์นี้ จะต้องกระทำโดยขบวนการที่เรียกว่า “การประเมินค่า” นั่นเอง การประเมินค่าที่ดี จำต้องอาศัยเครื่องมือวัด หลายชนิดที่สามารถให้ข้อเท็จจริงให้เราทราบว่า สถานะปัจจุบัน กำลัง เป็นอยู่เช่นไร และจะช่วยปรับปรุงสภาพเดิมนี้ให้ดีขึ้นได้ด้วยวิธีใดบ้าง

วัตถุประสงค์ของอนุสารฉบับนี้ ก็เพื่อจะช่วยให้ นักการศึกษา มอง เห็นคุณค่าของการวัดผลที่ดีให้ถูกต้องขึ้น พร้อมกับจะเสนอข้อแนะนำ บางประการที่จะปรับปรุงเทคนิค และเครื่องมือการประเมินค่าต่าง ๆ ให้ดีขึ้นด้วย ทั้งนี้โดยถือคติว่า การที่จะปรับปรุงการสอน การเรียน การแนะแนว การวางแผนหลักสูตร และประสิทธิภาพในการบริหารให้เจริญ พัฒนาขึ้นนั้น จำต้องอาศัยการวัดผลที่ดีและต่อเนื่องกันไปเป็นประการ สำคัญ คติดังกล่าวนี้ จะต้องเริ่มจากการที่นักศึกษามีทัศนคติที่ดีต่อการ

ประเมินผล พร้อมกับจะต้องไม่ยอมผ่อนปรน ต่อเครื่องมือวัดผลที่เลว และการสอบวัดที่ทำกันตามยถากรรมด้วย ผู้เขียนหวังเป็นอย่างยิ่งว่า อนุสารนี้คงจะลบล้างความคึกของครูอาจารย์บางท่านที่นึกว่า ข้อสอบที่ดี อาจจะสร้างขึ้นได้ง่าย ๆ โดยใช้เวลาเขียนเพียงไม่กี่นาทีก่อนที่จะถึงกำหนดสอบนั้นเสีย แต่ความจริงนั้น การทดสอบที่ดีก็เช่นเดียวกับการสอนที่ดี คือจะต้องมีทั้งการวางแผนและการเตรียมที่รอบคอบทุกประการ

การประเมินค่าและความมุ่งหมาย

หลักเบื้องต้นของขบวนการเรียน ก็คือการกำหนดความมุ่งหมาย ที่จะใช้เป็นปลายทางของการศึกษาและขบวนการที่จะนำไปสู่เป้าหมายนั้นๆ ความมุ่งหมายที่ดีจะต้องไม่คำนึงแต่เพียงสิ่งที่ครูสนใจจะสอนเท่านั้น แต่จะต้องคำนึงถึงสิ่งที่หวังปลูกฝังให้เกิดกับเด็กอีกด้านหนึ่งด้วย การตั้งความมุ่งหมายนี้มีความหมายยิ่งขึ้น ถ้าได้ให้นักเรียนมีส่วนร่วมรู้เห็นด้วย เพื่อเขาจะได้ทราบว่าครูหวังจะให้เขาสัมฤทธิ์ผลอะไรจากการเรียนนั้นๆ บ้าง และเมื่อเรียนจบบทเรียนแล้วหรือเมื่อเรียนไปถึงตอนใดตอนหนึ่งที่กำหนดไว้ ก็จะมีการประเมินค่ากันขึ้น เพื่อตรวจสอบดูว่าความมุ่งหมายต่าง ๆ ที่กำหนดไว้นั้นได้บรรลุผลเพียงใด ซึ่งโดยทั่วไปก็พอจะคาดได้ว่า ผลการเรียนนั้นจะไม่สอดคล้องกับความมุ่งหมายเหล่านั้นนัก ฉะนั้นเราจึงนิยมนำผลการสอบวัดด้านต่าง ๆ เหล่านี้มาวิเคราะห์ดูอีกที ว่านักเรียนสามารถสัมฤทธิ์ผลตามความมุ่งหมายเดิมที่ตั้งไว้ปานใด แต่ทั้งนี้มิใช่หมายความว่า เราจะต้องปรับปรุงความมุ่งหมายนั้นใหม่ โดยยึดเอาแต่ผลการเรียนของเด็กเป็นสำคัญ แต่หมายความว่า ความมุ่งหมายของการศึกษาที่แท้จริงนั้น จะต้องคำนึงแต่สิ่งที่นักเรียนสามารถประพฤติ และปฏิบัติตามได้เท่านั้น

ด้วยวิธีปรับปรุงแก้ไขไปตามลำดับเช่นนี้ ความมุ่งหมายเดิมที่เคยเขียนไว้ อย่างยืดยาวและเป็นอุดมคติเหล่านั้น ก็จะค่อย ๆ ลดน้อยลง จนกลายเป็น ความมุ่งหมายที่สั้นกว่าและจริงจังยิ่งขึ้น การปรับปรุงเหล่านี้ จะสำเร็จ ลงได้ก็ด้วยอาศัยการทดสอบที่ได้วางแผนไว้เป็นอย่างดีแล้วนั่นเอง

การประเมินค่าและการจูงใจนักเรียน

บทบาทที่สำคัญอีกประการหนึ่งของการประเมินค่า ก็คือการจูงใจ เด็ก ครูที่ฉลาดจะพยายามสนับสนุนให้นักเรียนประเมินค่าตนเอง ซึ่งเป็น วิธีที่จะเพิ่มความสนใจและการเรียนของเด็กให้กว้างขวางยิ่งขึ้น ซึ่งในเรื่อง นี้ การทดสอบในชั้นเรียนบ่อย ๆ ก็จะช่วยกระตุ้นให้เด็กกระตือรือร้นใน การเรียน และปรับปรุงงานของตนให้ดีขึ้นอยู่เสมอ และตามที่เราทราบ กันกันอยู่แล้วว่า ถ้านักเรียนรู้ตัวว่าเขาจะต้องถูกสอบไล่คัดสรรให้ได้ตก ตอนปลายปีเพียงครั้งเดียวแล้ว นักเรียนก็มักจะใช้วิธีทบทวน หรือกวาด วิชาอย่างลวก ๆ เฉพาะเมื่อใกล้จะสอบไล่เท่านั้น แต่ถ้านักเรียนทราบว่า เขาจะต้องถูกประเมินค่าผลการเรียนอยู่ตลอดเวลาแล้ว ก็จะทำให้เขาตั้งใจ เรียนและทุ่มเทแรงในการเรียนมากขึ้น เครื่องมือการสอบวัดและการ ประเมินค่าที่สร้างอย่างถูกต้องหลักวิชา จึงสามารถส่งเสริมเด็กให้เกิดนิสัยการ เรียนที่ดีได้เป็นอันมาก และโดยตรงข้าม ข้อสอบไล่ที่สร้างขึ้นอย่างผิด ๆ ก็จะทำให้นักเรียนสับสน บั่นทอนกำลังใจ และชักนำนักเรียนให้ผิดทาง อีกด้วย

การสอบและการให้เกรด

ความรับผิดชอบที่สำคัญอีกประการหนึ่งของครู ซึ่งถึงแม้ว่าจะ ยากแต่ก็จำต้องกระทำ ก็คือการให้เกรดนักเรียน ถ้านักเรียนถึงความสำคัญของ เกรดที่จะมีต่อนักเรียนแล้ว ครูก็จะต้องมั่นใจอย่างยิ่งว่า ตนมิได้ให้เกรด

นักเรียนตามอำเภอใจเลย เพราะเกรตจะเป็นตัวบอกทั้งผลสัมฤทธิ์ และคุณภาพในการเรียนของเขา และการที่นักเรียนได้เกรตสูงหรือต่ำนี้ จะเป็นตัวประกอบที่สำคัญที่จะช่วยให้นักเรียนสามารถได้งานทำ หรือได้รับอนุญาตให้เข้าสอบแข่งขันเข้ามหาวิทยาลัย หรือก่อให้เกิดความสนใจในแต่ละอาชีพ ตลอดจนจะเป็นแรงจูงใจหรือท้อใจต่อความสำเร็จในการเล่าเรียนอีกด้วย ฉะนั้นจะเห็นว่าการสอบไล่หรือการสอบวัดทั้งหลายนั้น จึงไม่เป็นแต่เพียงการบอกระดับผลสัมฤทธิ์ของตัวนักเรียนเองเท่านั้น แต่ยังบอกถึงผลการเรียนสัมพันธ์ระหว่างตัวนักเรียนกับเพื่อน ๆ ของเขาอีกด้วย ดังเหตุผลนานาประการดังกล่าวมานี้ เครื่องมือที่จะใช้ในการวัดผลและประเมินค่าทั้งปวง จึงจำต้องมีทั้งความเที่ยงตรงและความเชื่อมั่นให้มากที่สุด เพราะเครื่องมือเหล่านี้จะเป็นผู้ให้หน่วยการวัดที่ถูกต้อง ซึ่งจะนำไปใช้ในการพิจารณาให้เกรตต่อไปอีกด้วย

การประเมินค่าเป็นเครื่องมือของการเรียน

นอกจากการจูงใจแล้ว การประเมินค่ายังมีผลต่อขบวนการเรียนรู้ อีกด้วย คุณค่าของการสอบไล่ในวงการศึกษานั้น ย่อมขึ้นอยู่กับคุณภาพของการทดสอบโดยตรง การสอบที่ดีนับว่ามีคุณค่ามหันต์ — ส่วนการสอบที่เลวก็มีโทษอนันต์ดุจกัน การมีแบบทดสอบที่ดี และนำไปใช้อย่างถูกต้องนั้น จะสามารถช่วยให้เด็กแก้ไขข้อบกพร่องและความเข้าใจผิดของตนเองได้ด้วย และข้อสอบที่ดีก็ยังช่วยให้ทราบทั้งจุดเด่นและจุดด้อยของการเรียนและการสอนอีกโสดหนึ่งด้วย เมื่อใดมีความบกพร่องเกิดขึ้น ไม่ว่าจะเกิดกับเด็กนักเรียนบางคนหรือกับทั้งชั้นก็ตาม แบบทดสอบที่ดีก็จะช่วยให้เราทราบถึงสาเหตุของความบกพร่องนั้นได้เสมอ ยิ่ง

กว่านั้น ข้อสอบที่ดีก็ยังสามารถใช้เป็นอุปกรณ์การสอนได้อีก โดยครูนำข้อสอบนั้นไปเฉลยในชั้น ซึ่งจะทำให้นักเรียนทราบว่าครูมีความมุ่งหมายในการสอนการเรียนอะไรบ้าง และเขาควรจะนำประสบการณ์จากการสอบครั้งนั้นไปสัมพันธ์กับการสอนในบทเรียนอื่น ๆ ได้อย่างไรอีกด้วย

การประเมินค่าและการเข้าใจตัวเด็ก

ย่อมเป็นที่ทราบกันแน่ชัดแล้วว่า การศึกษามีใช่เป็นแต่เพียงการวางแผน และการสอนการเรียนเท่านั้น แต่โดยเนื้อแท้แล้ว หมายถึงขบวนการที่จะทำให้มนุษย์ฉลาดและสมบูรณ์ยิ่งขึ้น และเนื่องจากเรื่องนี้เป็นรากฐานที่สำคัญมาก จึงเป็นความจำเป็นที่ครูจะต้องจัด และเข้าใจซึ่งในเด็กแต่ละคนที่ตนสอนให้มากที่สุด ในปัจจุบัน เรามีเครื่องมือสำหรับวัดและประเมินผลนักเรียนมากมายหลายชนิด เช่น วัดความถนัด ทักษะ ทักษะการเห็นคุณค่า และวัดระดับความสามารถในการเรียน การปฏิบัติ และความสัมฤทธิ์ผลในด้านต่าง ๆ อีกด้วย ถ้าครูได้ทราบสมรรถภาพต่าง ๆ เหล่านี้ของนักเรียนแล้ว สถาบันการศึกษาก็จะสามารถวางแผนการศึกษาได้ถูกต้อง และกำหนดความมุ่งหมายที่แท้จริงได้สำเร็จ และยังสามารถใช้ผลการวัดต่าง ๆ เหล่านี้เป็นข้อมูลสำหรับการแนะแนว และให้คำปรึกษาแก่เด็กได้อย่างจริงจังอีกด้วย

สรุปความ

การประเมินผลการศึกษาที่ดี มีความหมายมากกว่าการสอบวัดความรู้อย่างธรรมดาๆ การประเมินผลที่ได้วางแผนอย่างถูกหลักวิชาและใช้เครื่องมือวัดที่เหมาะสมนั้น จะช่วยเพิ่มพูนให้ขบวนการศึกษาทั้งปวงสมบูรณ์ยิ่งขึ้น เช่น ช่วยในการกำหนดความมุ่งหมาย ก่อให้เกิดแรงจูงใจ

ในการเรียน นำให้เกิดความพัฒนาการในลักษณะต่าง ๆ ช่วยเสริมสร้างทัศนคติใหม่ ๆ ช่วยย่ำสิ่งที่เรียนมาแล้วให้ถาวร โดยนำความรู้ในวิชาการต่าง ๆ มาสัมพันธ์เข้าด้วยกัน ซึ่งจะทำให้เกิดความเข้าใจลึกซึ้งในวิชาการต่าง ๆ ยิ่งขึ้น และในที่สุด ก็จะเป็นรางวัลให้แก่นักเรียนที่เรียนที่เกิดความภูมิใจจากการที่ตนได้เกรดสูง การประเมินค่าที่คิมีได้ตรวจสอบแต่เพียงค่านักเรียนค่านเดียว แต่ยังช่วยให้ครูทราบข้อเท็จจริงต่าง ๆ จากผลการสอบวัดนั้น ๆ ไปปรับปรุงการสอนของตนให้ดียิ่งขึ้นต่อไปอีกด้วย

II ประเภทของเครื่องมือวัดผล

ดังที่กล่าวมาแล้วว่า งานที่ยากลำบากประการหนึ่งของครู ก็คือการประเมินความงอกงามของเด็ก และการประเมินความสามารถของตัวครูเอง เรื่องนี้จะค่อยง่ายขึ้น ถ้าครูได้รู้จักชนิดของเครื่องมือวัดแบบต่าง ๆ และทราบว่าชนิดไหนมีหน้าที่ในการวัดด้านใด ในสมัยนี้ นักวัดผลได้ประคิษฐ์เครื่องมือและวิธีการใหม่ ๆ สำหรับใช้วัดผลการศึกษาได้อย่างเที่ยงตรงถูกต้องเป็นอันมาก ดังเช่น เครื่องมือสำหรับวัดทักษะ ทัศนคติ คุณธรรมประเภทต่างๆ ฯลฯ ซึ่งสิ่งเหล่านี้ในสมัยก่อนเคยคิดว่า เป็นคุณธรรมที่ไม่สามารถจะวัดได้ ในอนุสารนี้ จะกล่าวแต่เพียงเครื่องมือวัดแบบสามัญ 3 ประเภทซึ่งใช้กันอยู่ทั่วไป และเป็นชนิดที่ไม่ต้องอาศัยเครื่องมือพิเศษอื่นใดมาประกอบ ซึ่งครูทั่ว ๆ ไปจะไม่สามารถจัดหาได้ เครื่องมือดังกล่าวนี้ก็คือ การสอบแบบปากเปล่า การสอบข้อเขียน (แบบให้ตอบเสรี) และการสอบแบบจำกัดคำตอบ

การสอบแบบปากเปล่า

ในบางกรณี การสอบปากเปล่านั้นก็เป็นวิธีวัดที่มีประสิทธิภาพมาก สามารถตรวจค้นให้ทราบว่า นักเรียนรู้อะไรมาแล้วบ้าง และเขาสามารถ

นำความรู้และประสบการณ์ ไปประสานสัมพันธ์กับความเข้าใจในด้านต่างๆ ได้ลึกซึ้งปานใด แต่เนื่องจากการสอบด้วยวิธีนี้สิ้นเปลืองเวลาและแรงงานมาก เขาจึงนิยมใช้วิธีสอบแบบอื่น ๆ แทน โดยเฉพาะในเมื่อมีนักเรียนที่จะต้องทดสอบเป็นจำนวนมาก อย่างไรก็ตามการสอบปากเปล่าก็ยังมีส่วนดีอยู่หลายประการ คือ (1) เป็นการสอบที่ทั้งครูผู้สอบและนักเรียนได้พบกันตัวต่อตัว ซึ่งจะช่วยให้นักเรียนเข้าใจความมุ่งหมายของคำถามได้ถูกต้องยิ่งขึ้น ทั้งนี้เพราะครูมักจะพบอยู่บ่อย ๆ ว่า เด็กเข้าใจความหมายของคำถามคลาดเคลื่อน เลยทำให้เขาตอบไม่ได้คะแนนดีทั้งๆ ที่เขามีความรู้ในเรื่องนั้นอย่างแจ่มแจ้ง (2) การสอบปากเปล่าสามารถช่วยให้ครูหยั่งและไล่เรียงความรู้ของนักเรียนเป็นเรื่องราวได้ว่าเขาเข้าใจในเรื่องนั้นลึกซึ้งปานใด—คือวัดความสามารถในการสัมพันธ์ ความคิดหลายๆ ด้าน เข้าด้วยกันได้หรือไม่ การถามปากเปล่าช่วยให้การสอบวัดเป็นไปอย่างกว้างขวางมาก (3) การสอบแบบนี้ช่วยให้ครูสังเกตปฏิกิริยาท่าทางของนักเรียนขณะที่ถูกสอบถามได้ด้วย เช่นนักเรียนมีการลังเล หรือความตื่นเต้นและตรึงเครียดปานใด ฯลฯ ซึ่งจากการสังเกตนี้ก็จะช่วยให้ครูสามารถประเมินความสามารถของเด็กได้ถูกต้องยิ่งขึ้น (4) ในบางกรณี การสอบปากเปล่านั้นก็เป็นสิ่งจำเป็นเลยทีเดียว เช่นในการสอบเด็กพิการ (ตาบอด หรือบังกับกล้ามเนื้อในการเขียนไม่ได้) ฯลฯ (5) นักเรียนบางคนสามารถปรับตัวเมื่ออยู่ต่อหน้าผู้อื่นได้เก่งกว่าธรรมดา ฉะนั้นผู้สอบที่ดีจะต้องระวังไม่นำเอาบุคลิกภาพของนักเรียน ไปปนกับความรู้ความสามารถของผู้สอบ (6) การสอบแบบปากเปล่านั้นดูที่จะสอดคล้องกับสภาพทั่วไปในชีวิตจริงเป็นอันมาก คือเด็กจะต้องพูดจาโต้ตอบกับผู้อื่นทั้งในด้านวิชาการ และเรื่องราวต่างๆ กันตัวต่อตัวเสมอ ฉะนั้นจึงเป็นวิธีสอบแบบธรรมชาติอยู่มาก เพราะตามปกติวิสัยนั้น บุคคลทั่วไปก็จะถามและตอบเรื่องราวต่างๆ โดยการพูด

มากกว่าที่จะโต้ตอบกันโดยการเขียน ฉะนั้นการสอบปากเปล่าจึงเสมือนกับเป็นการเพิ่มประสบการณ์ในด้านการสัมภาษณ์ให้แก่เด็กโดยตรงด้วยการสอบแบบปากเปล่านี้ น่าจะเป็นแบบที่เหมาะสมกับเด็กนักเรียน ที่ถูกสอนโดยวิธีตาม—ตอบ แบบธรรมดาๆ ทั้งที่พบกันทั่วไปเกือบทุกระดับการศึกษาในเมืองไทย เด็กที่เคยถูกสั่งสอนด้วยวิธีบรรยาย—โต้ถามจนเคยชินแล้ว จะรู้สึกว่าการสอบแบบนี้น่าสำหรับตนมาก

แม้ว่าการสอบแบบปากเปล่าจะมีส่วนที่อยู่หลายอย่าง— แต่ก็ยังมีมีส่วนเสียที่ฉกรรจ์อยู่หลายประการเช่นกัน เช่น (1) การทดสอบชนิดนี้มีความบกพร่องอย่างยิ่งตรงที่ ไม่มีความเชื่อมั่นเลย (คู่อภิธาน—ความเชื่อมั่น) ทั้งนี้เพราะเป็นการเหลือวิสัย ที่จะให้การตัดสินใจของครูที่มีต่อคำตอบของนักเรียนทุกคนคงที่และสม่ำเสมอ ถ้าการสอบนี้สอบด้วยกรรมการหลายคน วิธีที่ดีขึ้นก็โดยนำเอาผลการพิจารณาของกรรมการทุกคนมาร่วมวินิจฉัยให้เป็นคะแนนการสอบของเด็ก แบบเดียวกับการตัดสินใจผลการชกมวยตามความเห็นส่วนใหญ่ของกรรมการ แม้ว่าการใช้ความเห็นส่วนรวมนี้อาจจะไม่ถูกเสมอไปก็ตาม แต่ก็ยังนับว่าเป็นวิธีที่ให้ความยุติธรรมแก่เด็กอย่างดีที่สุดแล้ว (2) ข้อเสียประการที่สองของการสอบปากเปล่าทั่วไปก็คือการขาดความเที่ยงตรง (คู่อภิธาน) นั่นคือ การสอบนั้นมิได้สอบวัดสิ่งที่เราปรารถนาจะวัด ดังเช่นที่ได้กล่าวมาแล้วว่า กรรมการผู้สอบจะต้องระมัดระวังไม่หลงใหลต่อผู้สอบที่พูดคล่องแคล่ว (เว้นแต่ว่าเราต้องการจะสอบความคล่องแคล่วทางภาษา) แต่จะต้องยึดอยู่กับความรู้และความสามารถของผู้สอบเป็นสำคัญ ข้อบกพร่องที่คล้ายกันอีกประการหนึ่งก็คือ กรรมการมักจะเกิดความลำเอียงเนื่องจากรูปร่างหน้าตา และอาจปฏิกิริยาของนักเรียน ที่เข้าสอบ เพราะอาจเป็นไปได้ว่า เด็กผู้ชายที่

แข่งกร้าวมักจะขัดแย้งกรรมการ เลยเป็นเหตุให้เขาสอบตกเพราะบกพร่อง
 ในมรรยาท มากกว่าตกเพราะขาดความสามารถในวิชานั้นก็ได้ และโดย
 นัยตรงข้าม เด็กผู้หญิงที่หน้าตาดี ก็อาจจะได้คะแนนสูง ๆ จากกรรมการ
 ผู้ชาย เนื่องจากความงามของเธอ มากกว่าสอบได้เพราะความรู้ความเข้าใจ
 ในวิชานั้นก็ได้ (3) กรรมการที่สอบปากเปล่า มักจะรู้สึกเวลาที่ให้สำหรับ
 การสอบนั้นน้อยเกินไป จนเป็นการยากที่จะสอบถามคลุมเครือวิชาได้พอ
 ควรเพราะนักเรียนมักจะอธิบายคำถามบางข้อยืดเยื้อเกินไป และฝ่ายกรรม-
 การเอง ก็ชอบไล่ถามเด็กว่าเขาเข้าใจในเรื่องนั้นลึกซึ้งเพียงใด เลยไม่มี
 เวลาพอที่จะถามความรู้ต่าง ๆ ให้ครอบคลุมและกว้างขวางได้ ฉะนั้น
 เวลาที่ใช้ในการสอบปากเปล่า 1 ชั่วโมงจึงหมดไปอย่างรวดเร็ว จนกรรม
 การยังไม่ได้ข้อเท็จจริงเพียงพอที่จะวินิจฉัยว่า นักเรียนมีความรู้ความ
 สามารถเพียงใด (4) ข้อบกพร่องอีกประการหนึ่งที่ทำให้การสอบปากเปล่า
 ขาดความเที่ยงตรงก็คือ กรรมการมักจะมิได้วางแผนไว้ว่า จะถามอะไรตาม
 ลำดับก่อนหลัง แต่มักจะถามเรื่อยไปตามแต่ที่ตนนึกออก นี่ก็เป็นอีกเรื่อง
 หนึ่งที่จะต้องระลึกไว้เสมอ ๆ ว่าจะทดสอบเพื่อครอบคลุมอะไรบ้าง และ
 จะต้องกตเตรียมการสอบให้รอบคอบ เพื่อกันมิให้การถามออกนอกเรื่อง
 นอกราว และจะถามวัดแต่เพียงสาระสำคัญของวิชานั้นเท่านั้น (5) ข้อบก-
 พร่องสุดท้ายของการสอบปากเปล่า ก็คือการสิ้นเปลืองเวลาของครูอย่าง
 เหลือคณา หากชั้นใดมีนักเรียนมาก การที่จะจัดสรรเวลาสำหรับสอบ
 ปากเปล่าให้แก่เด็กแต่ละคน ๆ จนทั่วชั้นนั้น ก็ย่อมไม่ได้ผลคุ้มค่าเลย
 ที่เดียว และถ้าการวัดผลกลายเป็นภาระหนักเกินไปแล้ว— —ความกดดัน
 นี้ก็มักจะเสริมให้การสอบปากเปล่า กลายเป็นการสอบที่ขาดทั้งความเที่ยง
 ตรงและความเชื่อมั่น มากยิ่ง ๆ ขึ้น.

การสอบแบบข้อเขียน

การสอบแบบข้อเขียน เป็นวิธีที่ใช้วัดผลสัมฤทธิ์ทางการเรียน ของนักเรียนอย่างแพร่หลาย การที่วิธีสอบแบบนี้ได้เข้ามาแทนที่การสอบ แบบปากเปล่าในโรงเรียน ก็เนื่องจากที่เห็นประโยชน์ว่า เป็นการมอบ หมายงานอย่างเดียวกันให้กับเด็กทุก ๆ คนในชั้น และเป็นวิธีที่สามารถ สอบนักเรียนได้พร้อมกันหลายคนในเวลาเดียวกัน การสอบแบบข้อเขียนนี้ มีอยู่หลายชนิด และชนิดที่ใช้กันมากก็คือแบบเรียงความและแบบให้เขียน ตอบสั้น ๆ การสอบแบบนี้ ก็มีทั้งข้อดีและข้อเสียเกี่ยวกับการสอบ ปากเปล่า และถ้าได้กล่าวถึงเรื่องนี้เสียสักเล็กน้อย ก็อาจจะช่วยให้ครู ตัดสินใจได้ง่ายขึ้นว่า วิธีใดจึงจะเหมาะกับวัตถุประสงค์ของการสอนใน แต่ละสถานการณ์มากที่สุด

ข้อสอบแบบเรียงความแพร่หลายมากในประเทศไทย ครูทั่วไป มีความเห็นว่า การสอบแบบนี้ให้คำตอบที่ข้อสอบข้อเขียนประเภทอื่น ๆ ไม่สามารถจะทำได้ และการสอบแบบตอบสั้นนี้ทำให้เด็กตอบคำถามตาม ทำนองที่สื่อให้เห็นความคิดของเด็กเองด้วย และยิ่งกว่านั้นยังเป็นการสอบ วัดความสามารถในการรวบรวมความรู้ต่าง ๆ ให้เข้าเป็นระบบ และเป็น การวัดทักษะในการแสดงออกของนักเรียนอีกด้วย ครูที่นิยมข้อสอบแบบ นี้โต้แย้งข้อสอบประเภทปรนัยแบบเลือกอย่างถูกต้องว่า ข้อสอบแบบ เรียงความเป็นการให้เด็กเสนอคำตอบ มิใช่เพียงแต่ให้เด็กสังเกตคำตอบ ซึ่งก็หมายความว่า เป็นการสอบที่ลดการเตาของเด็กลงได้มากที่สุด คำถาม แบบเรียงความที่สร้างอย่างถูกต้องนั้น สามารถวัดสมรรถภาพในการเลือก เพ้นเนื้อหาใจความสำคัญของเนื้อเรื่อง ความสามารถในการสัมพันธ์เรื่อง เหล่านั้น และความสามารถในการผสมผสานให้เป็นเรื่องราวที่สมบูรณ์ ได้อีกด้วย

ข้อสอบแบบเรียงความก็มีข้อเสียทั้งเช่นข้อสอบอื่น ๆ เหมือนกัน คือมักจะเป็นเครื่องมือวัดที่เชื่อมั่นมิได้เลย เพราะคำตอบอาจแตกต่างกัน ได้หลายทาง และการให้คะแนนให้เกรตกก็ลำบากเป็นอย่างยิ่ง ข้อเสียอีกประการหนึ่งที่คล้าย ๆ กับข้อสอบแบบปากเปล่า ก็ตรงที่กรรมการอาจจะไปหลงให้ความสำคัญทางภาษาศิลป์ของคำตอบ มากกว่าทางความรู้ในเนื้อหาวิชา หากคำถามไม่เฉพาะเจาะจงจริง ๆ แล้ว คำตอบแบบเรียงความ ของเด็กแต่ละคนก็จะพลอยมีส่วนถูกบ้างไม่มากก็น้อยเสมอทุกข้อไป ปัญหา เรื่องกรรมการให้คะแนนไม่คงเส้นคงวาจะปรากฏอยู่ในข้อสอบแบบนี้เสมอ ได้มีการศึกษาค้นคว้าถึงเรื่องนี้มาแล้วมากมายซึ่งก็ปรากฏผลว่า การที่ให้ คะแนนแตกต่างกันนี้ ขึ้นอยู่กับผู้ตรวจให้คะแนนในเวลาไหน และเป็น การให้คะแนนแก่คำตอบฉบับแรกหรือฉบับสุดท้าย ดังในการทดลองที่ ให้นักการศึกษาผู้ทรงคุณวุฒิ 47 ท่านตรวจกระดาษคำตอบของนักเรียน 10 คน ซึ่งก็ปรากฏว่านักเรียนได้คะแนนทั้งตารางข้างล่างนี้*

คำตอบฉบับที่	เกรตกที่กรรมการให้				
	A	B	C	D	F
1	0	1	17	13	8
2	0	1	28	11	7
3	1	6	33	5	2
4	1	2	30	8	5
5	2	10	32	3	0
6	5	14	24	3	1
7	0	9	15	13	3
8	2	9	19	13	3
9	1	1	9	15	21
10	1	1	20	14	6

* ข้อมูลได้จาก Dr. Osmond E. Palmer, cited in *Evaluation in the Vasic College*, Edited by Paul Dressel, Harper & Bros., New York, N.Y.

ผู้อ่านจะเห็นได้ทันทีว่า กรรมการได้พิจารณาตัดสินให้คะแนนแก่นักเรียนแตกต่างกันมาก ดังเช่น นักเรียนคนที่ 8 นั้น กรรมการ 2 ท่านจะให้คะแนนระดับดีเยี่ยม (A) แต่อีก 3 ท่านเห็นว่าสอบตกใช้ไม่ได้เลย (F) ส่วนอีก 9 ท่านเห็นว่าอยู่ในเกณฑ์ดี (B) แต่อีก 13 ท่านเห็นว่ายังอ่อน (D) และมี 19 ท่านที่เห็นว่าพอใช้ได้ (C)

นี่เป็นปรากฏการณ์ที่น่าเศร้ายิ่ง เพราะแม้แต่ในวงคณะกรรมการที่ทรงคุณวุฒิ ก็ยังไม่เป็นที่ประจักษ์ว่า การให้เกรดให้คะแนนของข้อสอบแบบเรียงความจะเป็นที่เชื่อมั่นแน่นอนได้เลย จากความแตกต่างในการให้คะแนนแบบ ๆ นี้เอง ครูและนักการศึกษาจึงถือหลักปฏิบัติที่จะใช้กรรมการอย่างน้อย 2 ท่านตรวจให้คะแนนข้อสอบเรียงความแต่ละฉบับ แล้วถือคะแนนเฉลี่ยของ 2 ท่านนั้นเป็นคะแนนของเด็ก แต่ถึงกระนั้น ก็ยังมีการทดลองที่น่าคิดอยู่อีกว่า ในการให้ครูชุดเดียวกันให้คะแนนคำตอบฉบับเดียวกันซ้ำในเวลาต่างกันนั้น ผลการให้คะแนนก็ยังแตกต่างกันออกไปแบบ ๆ เดียวกับที่ให้กรรมการ 47 ท่านเป็นผู้ให้คะแนนนั้นเช่นกัน ครูที่ตรวจข้อสอบเรียงความมักจะรู้ตัวเองว่า ผลงานครั้งก่อน ๆ ของเด็ก ความเหนื่อยล้าในการตรวจ ลายมือ การเขียนผิดไวยากรณ์ และความบกพร่องต่าง ๆ เหล่านี้ มีอิทธิพลต่อการให้คะแนนข้อสอบแบบเรียงความ เป็นอันมาก

การสอบข้อเขียนแบบให้ตอบเสรีอีกชนิดหนึ่ง ก็คือแบบให้ตอบสั้น ๆ ซึ่งนักเรียนจะต้องเขียนตอบเพียงคำเดียว วลีเดียว หรือประโยคเดียว เติมลงในช่องว่างเป็นต้น ข้อสอบชนิดนี้ก็เป็นวิธีที่จะให้นักเรียนแสดงความเห็นของตนเองออกมาเหมือนกับในแบบเรียงความ มิใช่ให้

นักเรียนสังเกตแล้วจะเลือกตอบทั้งเช่นในแบบเลือกตอบ ข้อสอบแบบนี้
มักใช้ ในการสอบย่อย หรือใช้ร่วมกับข้อสอบแบบเรียงความที่ให้นักเรียน
เขียนตอบยาว ๆ มาแล้วหลายข้อ ข้อสอบชนิดนี้มีประโยชน์ในการทดสอบ
ความรู้ในเนื้อหาวิชาต่าง ๆ ได้ดี เช่น ให้บอกถึงเรื่องราว, สิ่งของ, ชื่อ,
วันที่ คำศัพท์ เป็นต้น ความลำบากอย่างหนึ่งของข้อสอบแบบนี้อยู่ที่ตรง
ที่ครูจะต้องคอยตรวจราคาคำตอบนานาชนิดของเด็ก และการให้เกรดให้
คะแนนก็เป็นอันน้อยอย่างยิ่ง ส่วนข้อดีที่เห็นได้ชัดอย่างหนึ่งก็คือ เรา
สามารถตรวจให้คะแนนได้เร็วกว่าแบบเรียงความเป็นอันมาก และใน
เวลาสอบที่เท่า ๆ กันนั้น เราก็สามารถถามได้มากกว่า ซึ่งทำให้การ
ทดสอบสามารถวัดได้อย่างกว้างขวางกว่า แม้ว่าจะไม่ลึกซึ้งเท่าก็ตาม.

ยังมีข้อทดสอบแบบจำกัดคำตอบอีกหลายชนิด ซึ่งจะไต่กล่าว
ละเอียดต่อไปในภายหลัง ในตอนนี้จะพิจารณาแต่เพียงข้อดีและข้อเสีย
ของแบบทดสอบชนิดต่าง ๆ เสียก่อน

แบบทดสอบชนิดจำกัดคำตอบ (หรือปรนัย) ที่ใช้กันอย่างแพร่
หลายก็คือ แบบเลือกตอบ ฝ่ายที่นิยมข้อสอบชนิดนี้อ้างว่า เป็นแบบชนิด
ของข้อสอบที่พัฒนาขึ้นมาเพื่อจะแก้ไขสิ่งบกพร่องต่าง ๆ ของข้อสอบ
ชนิดอื่น ๆ โดยเฉพาะอย่างยิ่งในเรื่องการให้คะแนนที่เชื่อมั่นได้สูงกว่า แต่
ฝ่ายตรงข้ามก็วิจารณ์ว่ายังมีข้อบกพร่องในด้านอื่น ๆ อยู่บ้าง ดังที่ไต่กล่าว
ไว้แล้วในเรื่องข้อสอบแบบเรียงความ แต่ข้อสอบแบบเลือกตอบก็มีความดี
อยู่มากทั้งเช่น (1) จะเห็นได้ชัดว่าข้อสอบปรนัยแบบเลือกตอบสามารถถาม
ได้กว้างขวางตลอดเนื้อหา เพราะเด็กสามารถตอบได้รวดเร็ว ทำให้ถามได้
หลายข้อ (2) แม้กับครูที่มีภาระและชั่วโมงการสอนมาก ข้อทดสอบ

ชนิดนี้ ก็ยังสามารถตรวจให้คะแนนได้อย่างถูกต้องแน่นอนโดยบุคคลทั่วไป เพียงแต่ครูมอบบุญแจกคำตอบให้กับผู้ตรวจคะแนน สำหรับไปเทียบกับคำตอบของนักเรียนว่าเขาตอบตรงตามนั้นหรือไม่ เท่านั้น (3) ข้อสอบปรนัยแบบนี้ยังมีข้อดีในทำนองที่นักเรียนทุกคนจะต้องแก้ปัญหาคำถามโดยวิธีเดียวกัน เพราะตัวคำตอบที่ให้เลือกนั้นไม่ผันแปรไปตามอำเภอใจหรือตามปฏิกิริยาของนักเรียนแต่ละคน (4) ข้อที่พิเศษของข้อสอบแบบเลือกตอบก็คือ การตรวจให้คะแนนของการสอบนั้น มีความเชื่อมั่นเป็นอย่างยิ่ง. ข้อสอบแบบนี้ยังสามารถช่วยให้ครูทราบค่าความยากและอำนาจจำแนกของคำถามแต่ละข้อได้อีกด้วย(คู่อภิธาน)ซึ่งก็เท่ากับสามารถหาความเชื่อมั่นและความเที่ยงตรงของข้อสอบทั้งฉบับนั้นได้ นั่นเอง

ข้อสอบแบบเลือกตอบก็ยังมีข้อเสียอยู่หลายอย่างเหมือนกัน คือ (1) ถ้าครูปรารถนาที่จะจำกัดความสามารถในการเขียนของนักเรียน หรือ कुछทักษะในการเรียบเรียงเรื่องราวต่าง ๆ หรือความสามารถในการตั้งสมมุติฐาน หรือให้สร้างสรรค์สังเคราะห์แล้ว ข้อสอบแบบนี้ก็มีความเหมาะสมน้อยกว่าแบบเรียงความ. (2) ในบางเนื้อหาวิชาที่ยังไม่เป็นที่ยอมรับเป็นข้อยุติ ว่าสิ่งใดถูกหรือผิดโดยต้องแท้แล้ว (เช่นในด้านสุนทรียภาพ) การสอบด้วยแบบอื่น ๆ ที่มีใช้ปรนัยมักจะให้ผลดีกว่า. (3) ฝ่ายตรงข้ามของข้อสอบแบบจำกัดคำตอบมักจะแย้งว่า นักเรียนมักจะสอบได้คะแนนดีกว่าที่เขาควรจะได้ เพราะนักเรียนมีโอกาสดูคำตอบจากตัวเลือกที่ให้ไว้ (4) ข้อเสียประการที่ 4 ที่มักจะกล่าวกันก็คือ ข้อสอบชนิดนี้กำหนดทางเลือกให้แก่นักเรียนเป็นจำนวนตายตัว ฉะนั้นเด็กจึงไม่สามารถแสดงสิ่งที่รู้จริง ๆ ออกมาได้อีกเลย

วัตถุประสงค์ของอนุสารนี้มีใช้ถือทางว่า ข้อสอบแบบใดดีที่สุด จากที่บรรยายที่แล้วมาก็พอจะแสดงให้เห็นว่า ข้อสอบแต่ละแบบ ก็มีทั้ง ข้อดีและข้อเสียทั้งนั้น จึงเป็นหน้าที่ของครูที่จะต้องเลือกเอาเองว่า ใน สถานการณ์นั้น ๆ ควรจะใช้เครื่องมือแบบไหนจึงจะเหมาะสมที่สุด ครูแต่ละ ท่านย่อมอยู่ในฐานะที่ดีที่สุดที่จะตัดสินว่า ข้อสอบแบบใดจึงจะเหมาะสม กับ ความปรารถนาและความต้องการของแต่ละชั้น ของแต่ละวิชานั้นๆ เมื่อครู ได้ทราบจุดเด่นจุดด้อยของข้อสอบแต่ละชนิดแล้ว ก็จะสามารถตัดสินใจได้ อย่างเหมาะสมยิ่งขึ้น

III การสร้างข้อสอบ

ไม่ว่าครูจะเลือกใช้ข้อสอบแบบใดก็ตาม ก็ยังมีหลักการและคุณ ลักษณะต่าง ๆ ที่ร่วมกันอยู่หลายประการดังที่ได้กล่าวไว้ในตอนก่อนแล้ว ว่า การสอบนั้นมีหน้าที่อยู่หลายประการ ความมุ่งหมายประการแรกของ การสอบก็เพื่อจะทราบผลสัมฤทธิ์ของนักเรียน และความสามารถในการ นำสิ่งที่เรียนรู้แล้วไปใช้ ความมุ่งหมายประการที่สอง ก็เพื่อจะใช้ผลการ ทดสอบไปปรับปรุงการสอนการเรียนในอนาคตให้เกิดผลดียิ่งๆ ขึ้น ครูที่ ออกข้อสอบ จะต้องเป็นผู้มีความรู้ในเนื้อหาวิชานั้นๆ เป็นอย่างดี จึงจะ สามารถถามปัญหาแต่ที่มีสาระสำคัญ ๆ และสามารถตรวจคำตอบได้อย่าง ถูกต้อง คำถามทุกข้อจะต้องถามอย่างแจ่มแจ้ง เพื่อนักเรียนจะได้ทราบ แน่ชัดว่า ครูต้องการจะให้เขาตอบกว้างแคบปานใด ฝ่ายผู้ออกข้อสอบ ก็จะต้องตัดสินใจได้ว่า ในสภาพนั้น ๆ ควรจะออกข้อสอบแบบไหนจึงจะ เหมาะสม ฉะนั้น ก่อนที่จะเริ่มต้นเขียนข้อสอบจริง ๆ ครูจึงต้องคำนึงทั้ง ด้านความมุ่งหมายที่หลักสูตรกำหนดให้นักเรียนสัมฤทธิ์ และด้านผลการ เรียนที่จะเป็นเครื่องชี้บอกว่านักเรียนได้สัมฤทธิ์อะไรไปแล้วบ้าง การที่

ครูจะเลือกใช้ข้อสอบแบบใดนั้น ขึ้นอยู่เป็นบางประการกับวัตถุประสงค์ที่ครูต้องการจะวัดด้วย ข้อสอบที่ดีจะไม่วัดแค่เพียงว่า นักเรียนได้เรียนรู้อะไรไปแล้วบ้าง แต่จะต้องวัดความสามารถในการนำเอาความรู้และทักษะ อันเป็นผลจากการเรียนนั้นไปใช้ได้หรือไม่ อีกประการหนึ่งช่วยในการสร้างข้อสอบในขั้นต่อไป ครูควรจะได้เขียนหัวข้อของเนื้อหาวิชาและสมรรถภาพด้านความเข้าใจและอื่นๆ ตามที่ที่ต้องการจะวัดลงไว้ให้แน่ชัด ในการนี้ครูควรจากร่างโครงสร้างของข้อสอบพอเป็นเค้าขึ้นเสียขั้นหนึ่งก่อน แล้วนำไปให้เพื่อนร่วมงานช่วยตรวจทานและปรับปรุงตามควร ก่อนที่จะนำไปใช้สอบกับนักเรียนจริงๆ ตอนนี้นักครูจะต้องคำนึงถึงเวลาที่จะใช้ในการสอบและตัดสินใจว่า จะถามกี่ข้อและจะใช้คำถามแบบใด ซึ่งจะต้องพิจารณาถึงความยากของคำถามด้วย ถ้าครูต้องการเพียงจะสอบความทรงจำในเนื้อหาวิชา ก็ควรใช้คำถามที่ให้นักเรียนระลึกตอบเอาเอง ถ้าครูสอนบทเรียนนั้นดีแล้วและนักเรียนก็ตั้งใจเรียนดีด้วย คะแนนจากการสอบก็จะสูงมาก ถ้าในการสอบคราวอื่นที่ครูต้องการสอบเพื่อจำแนกเด็กในชั้น ว่าใครจะสัมฤทธิ์ผลในการเรียนมากน้อยต่างกันอย่างไรแล้ว คำถามของข้อสอบนั้นจะต้องมีทั้งง่ายมากๆ และบางข้อจะต้องยากขึ้นจนเด็กส่วนใหญ่ตอบถูก แต่เด็กอ่อนๆ ตอบผิด และจะต้องมีบางข้อที่ยากมาก ๆ อีกด้วย คือ เด็กเก่งมากๆ เท่านั้นจึงจะตอบข้อนั้นถูก เรื่องนี้จะได้กล่าวต่อไปในวิธีหาค่าความยากและอำนาจจำแนกของข้อทดสอบ ขณะนี้จะบรรยายถึงกลวิธีในการสร้างข้อสอบแบบต่าง ๆ พร้อมกับคำแนะนำวิธีใช้และคุณค่าของการใช้ข้อสอบแบบต่าง ๆ เหล่านั้นเสียก่อน

การสอบปากเปล่า

เราอาจใช้การสอบปากเปล่าได้หลายโอกาส ซึ่งก็แล้วแต่เนื้อหาวิชาและความมุ่งหมายเฉพาะของการทดสอบนั้น ๆ เช่นถ้าเราต้องการจะสอบทักษะในวิชาใด ๆ ของนักเรียน เราก็ใช้คำถามสั้น ๆ แต่มา ๆ ข้อ หรือเราอาจจะถามเน้นเฉพาะตรงตอนใดตอนหนึ่งของเนื้อหา นั้นเพื่อคือนักเรียนมีความเข้าใจในตอนนั้นลึกซึ้งปานใดก็ได้ ถ้าในโอกาสอื่นเราต้องการจะทดสอบการแสดงปฏิกิริยาของนักเรียนภายใต้สภาพความคับแค้นใจ เราก็ต้องเริ่มต้นโดยสร้างสถานการณ์อย่างใดอย่างหนึ่งให้นักเรียนโกรธเสียก่อน แล้วครูกอยชักคอทุกคำพูดที่เด็กตอบเป็นต้น ที่กล่าวมานี้มิได้ประสงค์จะแนะให้ครูที่จะสอบวัดความรู้เด็กนำไปปฏิบัติ แต่ต้องการที่จะชี้ให้เห็นว่า การทดสอบที่ดีนั้น จะต้องการผู้สอบที่มีความชำนาญจึงจะสร้างข้อสอบชนิดนั้นได้เที่ยงตรง ในการสอบปากเปล่าทั่วไปกรรมการสอบมักจะใช้ไม่วิธีอุปนัย (ถามจากส่วนย่อยไปหาส่วนใหญ่) ก็ใช้วิธีนิรนัย (ถามจากส่วนใหญ่ไปหาส่วนย่อย) การสอบที่มีการเตรียมไว้ล่วงหน้าย่อมดีกว่าการถามปะปะโดยปราศจากความมุ่งหมายอย่างแน่นอน ถ้าการถามได้ดำเนินไปตามระบบและตามลำดับก่อนหลังแล้ว อาจจะช่วยให้ให้นักเรียนระลึกถึงเรื่องราวต่าง ๆ ที่เรียนมาได้ดีขึ้นด้วย ข้อควรจำที่มีค่ายิ่งก็คือ ถ้าวัตถุประสงค์ของการสอบนั้น เพื่อจะวัดความรู้ที่นักเรียนแล้ว ก็ควรจะทำให้สิ่งแวดล้อมเหมาะสมเสมอ ซึ่งก็หมายความว่านักเรียนจะต้องไม่รู้สึกรู้สึกระหนกใด ๆ เลย แต่ควรจะทำให้เขารู้สึกเป็นกันเองตามสบายเป็นต้น ถ้ากรรมการสอบปากเปล่าจะโอภาปราศรัยก่อนเวลาสอบจริงสัก 2-3 นาที ก็จะช่วยให้มีการสอบนั้นลุล่วงด้วยดียิ่งขึ้น

ในการสอบปากเปล่า ครูจะต้องเตรียมการและคิดให้รอบคอบเสียก่อนว่า ต้องการจะสอบวัดถึงเรื่องอะไรบ้าง นั่นคือ จะถามอะไร และจะประเมินให้คะแนนคำตอบเหล่านั้นอย่างไรด้วย ครูสอบควรจะมีหัวข้อที่จะถามไว้ในมือ และกำหนดน้ำหนักคะแนนของแต่ละหัวข้อนั้นไว้ด้วยว่าตอนไหนควรจะให้คะแนนเท่าไร และในขณะที่ทำการสอบอยู่นั้น ครูสอบก็ต้องมีวิธีการสำหรับบันทึกผลการสอบของแต่ละตอนนั้นไว้ด้วย ถ้ามีความต้องการหรือจำเป็นที่จะต้องบันทึกผลการสอบไว้ให้ถาวร (หรือครูต้องการจะทบทวนคำตอบนักเรียนอีกในโอกาสต่อไป) ก็ควรจะใช้เครื่องบันทึกเสียงบันทึกคำตอบนั้นไว้เลย การสอบและบันทึกแบบนี้จะมีคุณค่ายิ่งขึ้น ถ้าจะได้นำผลการสอบปากเปล่าครั้งนี้ ไปเปรียบเทียบกับผลการสอบแบบข้อเขียนของนักเรียนหรือ กับผลงานในชั้นของเขา ฯลฯ ถ้าปรากฏว่าผลจากการสอบปากเปล่าแตกต่างกับผลการสอบแบบอื่น ๆ ของนักเรียนมากมาย เราก็ควรที่จะกลับไปพิจารณาถึงประสิทธิภาพของการสอบเหล่านั้นอีกที หรืออาจจะเรียกเด็กมาไต่ถามเพิ่มเติมอีกก็ได้

เหนือสิ่งอื่นใด ทั้งครูและผู้สอบจะต้องระลึกถึงวัตถุประสงค์ของการสอบไว้เสมอ ๆ ถ้าผู้สอบดำเนินการใดๆ ที่นอกเหนือไปจากการวัดความรู้ของเด็กแล้ว ก็อาจจะทำให้เด็กหลงผิดไม่แสดงความสามารถสูงสุดของเขาออกมาอย่างเต็มที่ เรื่องนี้หมายความว่า ผู้ดำเนินการสอบควรจะต้องหลีกเลี่ยงการเสริมคำตอบให้แก่เด็กโดยทุกวิถีทาง นอกจากจะกระทำเพื่อสะกิดให้นักเรียนหวนระลึกถึงเรื่องที่เขากำลังนึกไม่ออกในขณะนั้นเท่านั้น ข้อบกพร่องที่ร้ายแรงอีกประการหนึ่งที่ครูสอบควรระวังไว้เสมอก็คือ ไม่ควรจะต้องถามแต่เพียงเรื่องใดเรื่องหนึ่งมากเกินไป ซึ่งจะกลายเป็นการวัดความลึกซึ้งของเรื่องนั้นเพียงเรื่องเดียว โดยมีได้วัดความรู้ในด้านอื่น ๆ ให้กว้างขวาง

การสอบข้อเขียนแบบเรียงความ

ในตอนก่อนได้กล่าวถึงข้อเสียของการสอบแบบเรียงความ โดยเน้นถึงความบกพร่องในการให้เกรคหรือให้คะแนน แต่ทั้งนี้ก็ได้หมายความว่าควรจะเลิกใช้ข้อสอบแบบเรียงความนั้นเสียเลย แต่โดยตรงกันข้าม ข้อสอบแบบเรียงความก็ยังเป็นเครื่องมือที่มีประสิทธิภาพอยู่มาก โดยเฉพาะในวิชาที่มุ่งหมายจะวัดความสามารถในการรวบรวมความคิด หรือที่ต้องการให้นักเรียนสร้างสรรค์การเขียนขึ้นเอง หรือวัดความสามารถที่จะสัมพันธ์ความคิดกันต่าง ๆ เข้าด้วยกัน ฯลฯ จุดอ่อนของข้อสอบแบบเรียงความมักเกิดจากการเขียนโจทก์คำถามเป็นส่วนใหญ่ คือคำถามนั้นมักจะถามเสียกว้างขวางและครอบคลุมมากเกินไป จนเด็กไม่ทราบชัดว่าครูต้องการจะถามอะไรกันแน่ หรือไม่อีกที ก็ถามจำกัดเสียแคบเกินไป จนแทบจะกลายเป็นคำถามแบบให้เติมคำในช่องว่างเสียเลยทีเดียว ตามที่ถูกที่ควรนั้น คำถามแบบเรียงความควรจะแจ่มชัดพอที่เด็กจะเข้าใจได้ถูกต้องว่า ครูต้องการจะถามอะไร และควรมีคำชี้แจงพอที่จะแนะนำให้นักเรียนทราบถึงเค้าโครงที่จะตอบด้วย

ในการตั้งคำถามแบบเรียงความนั้น ควรจะเป็นคำถามที่วัดความรู้จนครอบคลุมทั่วเนื้อหาที่สอน ไม่ควรถามรายละเอียดแต่เพียงตอนใดตอนหนึ่งจากเนื้อหาที่สอนไว้ตลอดภาคเรียนของวิชานั้นเลย การถามที่ครอบคลุมจะช่วยให้ข้อสอบมีความเชื่อมั่นได้ยิ่งขึ้น เพราะเป็นการสอบวัดผลสัมฤทธิ์เบ็ดเสร็จของเด็กทุก ๆ ท่าน มิใช่จำกัดเฉพาะเจาะจงแต่เพียงตอนหนึ่งตอนใดเท่านั้น ฉะนั้นการสอบชนิดนี้ จึงต้องให้เด็กตอบคำถามหมดทุกข้อ โดยไม่มีการเลือก วิธีที่จะออกข้อสอบได้ดี ก็โดยเริ่มจากการเขียนตารางกว้างขวาง โดยให้ด้านหนึ่งเป็นเนื้อหาวิชา และอีก

ด้านหนึ่ง เป็นทักษะ หรือพฤติกรรม ที่เรา ประารถนา จะให้เกิดขึ้น ในตัวเด็ก
ทั้งนี้

เนื้อหาวิชา	สมัยกรุงศรี อยุธยา	สมัยรัตนโก- สินทร์รัชกาล ที่ 1-4	สมัยรัตนโก- สินทร์รัชกาล ที่ 5-7	สมัยปฏิวัติ ปี พ.ศ. 2475 ถึงปัจจุบัน	รวม
ความรู้	1	2	1	2	6
ความเข้าใจ	3	1	2	1	7
การนำไปใช้	1	1	2	3	7
วิจารณ์ญาณ	2	1	1	1	5
รวม	7	5	6	8	26

จากตารางนี้เราก็จะทราบได้ว่า ในการสอนควรจะเน้นในเรื่องใด
มากที่สุด และในการสอบควรจะถามหนักไปในด้านใดบ้าง และถ้าเรา
ได้วิเคราะห์จำนวนและชนิดของข้อคำถามต่าง ๆ ในข้อสอบที่จะใช้ในการ
ทดสอบนั้นแล้ว เราก็จะทราบได้ว่า การสอนในเทอมที่แล้วของเรานั้น
ครอบคลุมเนื้อหาอะไรบ้างและทั่วถึงปานใด

ก่อนที่จะทำตารางนี้ ครูจะต้องกำหนดความมุ่งหมายเบื้องต้น
ของวิชานั้นให้แน่ชัดเสียก่อน เช่นถ้าความมุ่งหมายประการแรกต้องการ
ให้เด็กมีความรู้ เราต้องบ่งให้ชัดลงไปว่าเป็นความรู้ชนิดใด เป็นความรู้
ในเนื้อเรื่องหรือเป็นความรู้รวมยอด หรือเป็นความรู้ในการขยายหลัก
วิชา ฯลฯ หรือถ้าวิชานั้นมีความมุ่งหมายที่จะให้เด็กเกิดความเข้าใจ
ก็หมายถึงความเข้าใจในการแปลความหรือตีความ แม้ความมุ่งหมายอื่นๆ
เช่น การนำไปใช้ วิจารณ์ญาณ และ การประเมินค่า เหล่านี้ ก็ควร
ต้องบ่งให้ชัดเจนแบบเดียวกันหมด และไม่ว่าจะมีความมุ่งหมายอื่นใด
อีกก็ตาม ครูก็จะต้องคำนึงถึงสิ่งเหล่านั้นในการเขียนข้อสอบเสมอ ทั้งนี้
ก็เพื่อจะได้ทราบว่า การสอนของเราได้บรรลุผลเพียงใด

ในการเขียนข้อคำถามนั้น ครูจะต้องระมัดระวังในการใช้ถ้อยคำ เพื่อให้เด็กทราบแน่ชัดว่าเขาจะต้องตอบอย่างไร ถ้าจะเขียนคำถามว่า “จงบรรยายเรื่องการปกครองแบบสมบูรณาญาสิทธิราช มาตามที่ท่านทราบ” หรือ “จงอภิปรายถึงการพัฒนาการของวิธีการวิทยาศาสตร์” เช่นนี้แล้ว เด็กก็อาจเขียนตอบได้สารพัดชนิดโดยต่างก็มีส่วนถูกด้วยกันทั้งนั้น นี่คือคำถามแบบครอบจักรวาลที่กว้างขวางเกินไป แต่ถ้าถามว่า “จงบรรยายถึงนโยบายต่างประเทศของรัชกาลที่ 5 สมัยที่ต้องเผชิญกับลัทธิล่าอาณานิคมของฝรั่งเศส” เช่นนี้แล้ว ก็จะเป็นการกำหนดประเด็นของคำถามให้จำเพาะเจาะจงยิ่งขึ้น ทำให้การประเมินคำตอบเชื่อมั่นได้มากยิ่งขึ้นด้วย มีคำถามบางคำถามที่จะช่วยให้คำตอบของนักเรียนจำเพาะยิ่งขึ้น เช่น “จงให้คำจำกัดความของ.....จงเลือก.....จงสรุป.....จงเปรียบเทียบ..... ฯลฯ” คำถามเหล่านี้จะช่วยสัมพันธ์เนื้อหาวิชาเข้ากับความมุ่งหมายได้เป็นอย่างดี ซึ่งเป็นสิ่งสำคัญที่ครูจะต้องคำนึงอยู่ตลอดเวลาที่เขียนข้อสอบ

จงพิจารณาคำถามประเภทเรียงความต่อไปนี้ แล้วลองคิดว่า เราควรจะปรับปรุงคำถามเหล่านี้อย่างไร จึงจะสามารถใช้วัดผลสัมฤทธิ์ในการเรียนของเด็กตามความมุ่งหมายนั้นได้ ประการแรกคือ **คำถามสองแง่** ที่ถามว่า “จงบรรยายถึงกำเนิดประชาธิปไตยของประเทศไทย” คำถามชนิดนี้มีได้บอกแนวทางอะไรแก่เด็กเลย คำว่า “จงบรรยาย” นั้นหมายถึงอะไร ? จะหมายถึงให้นักเรียนบอกรายชื่อบุคคลที่ส่งเสริมขบวนการประชาธิปไตย หรือจะให้นักเรียนกล่าวถึงองค์ประกอบที่สำคัญที่นำไปให้เกิดลัทธินี้ ? และคำว่า “กำเนิด” นั้นเล่าหมายถึงอะไรกันแน่ ? จะให้นักเรียนตอบตั้งแต่ความไผ่ผ่นของพระยาพหล หรือจะให้ท้าวความถอยหลังไปถึงพระราชดำริห์ของ พระมหากษัตริย์ ที่ทรงพระมหากรุณา ต่อการ

ปฏิวัติใน พ.ศ. 2475? และคำ “ประชาธิปไตย” นั้นก็หมายถึงอะไรกัน - ประชาธิปไตยการเมือง? สังคมประชาธิปไตย? หรือประชาธิปไตย เศรษฐกิจ? ฯลฯ คำถามเช่นนี้อาจปรับปรุงเสียใหม่จนนักเรียนต้องตอบไปตามแนวเดียวกันได้ ดังเช่น “จงสรุปคุณงามความดีของ พระยาพหลที่มีต่อลัทธิประชาธิปไตยในปัจจุบัน” หรือ “พระยาพหลมีความคิดเห็นในด้านการเมืองแตกต่างกับผู้นำประเทศในสมัยก่อน ๆ นั้นอย่างไร?” เป็นต้น

คำถามคลุมเครือ เช่น “จงอภิปรายถึงผลของระบอบประชาธิปไตย” คำถามข้อนี้อาจปรับปรุงให้จำเพาะเจาะจงยิ่งขึ้นดังเช่น “จงบรรยายถึง ภัยพิบัติที่บังเกิดขึ้นในทันทีที่ลัทธิระบอบประชาธิปไตยตกที่ ชิโรชิม่า และผลต่างๆ ที่เกิดขึ้นในกาลต่อมาหลังจากการทิ้งระเบิดนั้น ในการบรรยายจงกล่าวถึง ผลที่มีต่ออาคารบ้านเรือน ประชาชน และการเปลี่ยนแปลงของทัศนคติที่มี ต่ออาวุธปรมาณูด้วย”

ถ้าจะจำแนกคำถามออกเป็นประเภท ๆ ตามความมุ่งหมายของ หลักสูตรพร้อมกับยกตัวอย่างประกอบแล้ว ก็อาจช่วยให้เราเห็นง่ายขึ้น ดังนี้

ประเภทวัดความรู้

1. แบบระลึกสามัญ

“ใครเป็นผู้นำฝ่ายไทยในการทำศึกค่ายบางระจัน?”

2. แบบระลึกจำเพาะ

“ปัญหาระหว่างประเทศที่สำคัญ 4 ประการ ในปี ค.ศ. 1890-1913 ที่ทำให้เกิดสงครามโลกครั้งที่ 1 นั้น คืออะไร?”

ประเภทวัดความเข้าใจ

1. แบบให้เปรียบเทียบ

“ร่างกายของสัตว์ที่เลี้ยงลูกด้วยนม กับสัตว์ครึ่งน้ำครึ่งบก แตกต่างกันตรงไหนบ้าง?”

“การปกครองแบบ ประชาธิปไตย กับแบบ สาธารณรัฐ ต่างกันอย่างไร?”

2. แบบให้อธิบายความหมายของวลีหรือข้อความ

อตุตถอนุตา หมายถึงอะไร?

คำสอนที่กล่าวนี้หมายถึงอะไร “พึงอย่าฝั่งทรัพย์ไว้ในดิน ซึ่งอาจถูกทำลายโดยปลวกและสนิมหรือโจรภัย แต่พึงฝังไว้ในสวรรค์ ซึ่งทั้งปลวกและสนิมมิอาจทำลายได้ และโจรก็มีอาจขโมยไปไม่ได้ด้วย”

3. แบบให้สรุปเรื่องราวหรือสังเคราะห์ข้อความ

“การจัดสรรโดยธรรมชาติ” ของคาร์วิน นั้นหมายถึงอะไร?

ประเภทนำไปใช้

ถ้าเราต้องการสอบวัดความรู้ว่า นักเรียนสามารถนำความรู้ของเขาไปใช้ในการแก้ปัญหาใหม่อื่น ๆ ที่นอกเหนือไปจากสิ่งที่เคยรู้เคยเข้าใจมาแล้ว เราก็มีวิธีถามได้หลายอย่าง แต่ตรงนี้จะสังเกตเห็นได้ว่า คำถามประเภทวัดความเข้าใจ กับประเภทนำไปใช้นั้นคล้ายคลึงกันมาก ความเข้าใจหมายถึงความสามารถที่จะระลึกออก และอธิบายเรื่องราวหรือหลักการใด ๆ ได้ ส่วนการนำไปใช้หมายถึงการนำความรู้และความเข้าใจมารวมกัน จนสามารถนำไปแก้ปัญหาใหม่ที่นักเรียนยังไม่เคยทราบวิธีคิดมาก่อน เรามีวิธีทดสอบความสามารถด้านการนำไปใช้ได้หลายแบบ ดังนี้

1. วัดความสามารถในการย่อและขยายปัญหาต่างๆ ในชีวิตจริง
“การที่ ประเทศไทย มี นโยบายที่จะขยายการศึกษาประเภท
โรงเรียนมัธยมแบบประสมออกไปให้กว้างขวางนั้น จะทำ
ให้การสอบการวัดผลเปลี่ยนแปลงไปอย่างไรบ้าง ?”
2. การวิเคราะห์และเปรียบเทียบ
“เหตุใด ในสมัยปัจจุบัน ครูจึงนิยม การทดสอบแบบข้อเขียน
มากกว่าแบบปากเปล่า ?”
3. ความสามารถในการนำความคิดเห็น และ/หรือความกึ
รวบยอดจากสถานการณ์หนึ่งไปสัมพันธ์กับสถานการณ์อื่น ๆ
“ศาสนาพุทธมีอิทธิพลต่อระบบการสอนและหลักสูตร อย่าง
ไรบ้าง ?”
4. การให้ยกตัวอย่างตามหลักวิชา
“เราจะใช้ขบวนการ วิทยาศาสตร์ ไปพิสูจน์ว่า นมบูด เพราะ
แบคทีเรียโดยวิธีใด ?”

ประเภทวิจารณ์ญาณ

มีความมุ่งหมายอีกประการหนึ่งที่มีมักจะปรากฏอยู่ในบทเรียนต่างๆ
เสมอ สิ่งนี้ก็คือความสามารถในการประเมินคุณค่า หรือตัดสินชี้ขาด
ปัญหาต่าง ๆ นักเรียนจะถูกสอนให้รู้จักวิธีสำรวจคุณค่า, ความดี, ความ
แม่นยำ, ความลำเอียง ฯลฯ ของเรื่องราวต่าง ๆ ความสามารถในการ
ตัดสินชี้ขาด เป็นสิ่ง สำคัญ ที่จะทำให้เกิดความเข้าใจในเนื้อหาวิชานั้น
อย่างลึกซึ้ง การตัดสินชี้ขาดนี้ ไม่เหมือนกับความเห็น ซึ่งเป็นแต่เพียงการ
ตัดสินใจที่ปราศจากข้อสนับสนุนที่มั่นคง เราสามารถตั้งคำถามประเภท
วิจารณ์ญาณนี้ได้หลายแบบเหมือนกัน ดังเช่น

1. แบบให้ตัดสินว่าถูกหรือผิด— —สนับสนุนหรือคัดค้าน
 “ที่กล่าวว่า ‘มนุษย์เป็นเกณฑ์สำหรับวัด’ นั้นมีความจริง
 ปานใด ?”
 “ควรเลิกระบบการสอบคัดเลือกในวงการศึกษาหรือไม่เพราะ
 เหตุใด ?”
2. แบบให้ตรวจสอบความแม่นยำและลำเอียง
 “ท่านจะชี้ให้เห็นได้อย่างไรว่า ผู้แต่งตำราเล่มนี้ ไม่นิยมเล่น
 การเมืองเลย ทั้ง ๆ ที่เขาก็มีได้กล่าวเอาไว้อย่างเปิดเผย ?”
 “ท่านมีความเห็นเช่นไรต่อกำกล่าวที่ว่า รัฐบุรุษเยอรมัน
 คำหนิอังกฤษว่าเป็นผู้ก่อสงครามโลกครั้งที่ 1”

การตรวจให้คะแนนข้อสอบแบบเรียงความจกว่าเป็นงานยากที่นำ
 เบื่ออย่างหนึ่งของครู ทั้งนี้ก็เนื่องจากความเป็นอัตนัยในคำตอบที่นัก-
 เรียนเขียนตอบ ผสมกับอัตนัยจากการอ่านของครูประกอบกัน วิธีที่จะ
 ช่วยให้การตรวจให้คะแนนมีความเชื่อมั่นได้สูงขึ้นนั้น จะต้องระลึกถึงหัว
 ข้อ 2—3 ประการดังนี้ ถ้าข้อสอบนั้นมีหลายข้อ ก็ควรตรวจให้คะแนนทีละ
 ข้อ ๆ จนตลอดหมดทุกฉบับ เหตุผลของเรื่องนี้เห็นได้ง่าย คือถ้าครูได้
 ตรวจคำตอบแรกของฉบับใดและให้คะแนนไปเท่าใดแล้ว คะแนนของข้อ
 แรกนั้นก็มักจะมียุทธพลสืบเนื่องไปถึงคะแนนของข้อหลังๆ ด้วยเสมอ และ
 ประการที่สองนั้น ถ้าครูไม่ได้อ่านคำตอบข้อเดียวกันของนักเรียนทุกคน
 พร้อมกันแล้ว ก็ไม่สามารถเปรียบเทียบคุณภาพการตอบของเด็กได้อย่าง
 ถูกต้องด้วย บางทีครูอาจใช้วิธีอ่านกระดาษคำตอบของแต่ละฉบับจนตลอด
 หมดทุกข้อ แล้วแบ่งออกเป็น 3 หรือ 4 กลุ่มก่อน ตามประเภทดีมาก-ดี-พอใช้-
 —อ่อน หรือยังไม่พอใช้ก็ได้ แต่ถ้การตรวจนั้นจะต้องให้คะแนนหรือให้

เกรกแล้ว ครูก็จะต้องนำคำตอบของแต่ละกลุ่มนั้น ๆ มาอ่านซ้ำเพื่อเปรียบเทียบกันอีกทีหนึ่ง จึงค่อยให้คะแนน และเพื่อให้คะแนนเหล่านี้มีความเป็นปรนัยยิ่งขึ้น ครูผู้ตรวจต้องไม่ควรดูชื่อนักเรียนเลย เพราะการรู้ชื่ออาจมีอิทธิพลต่อเกรกมากกว่าคำตอบก็ได้

ในการตรวจสอบประเภทเรียงความนั้น มีวิธีการอยู่อย่างหนึ่งที่จะช่วยให้การให้คะแนนของครูสม่ำเสมอยิ่งขึ้น โดยครูเฉลยสิ่งที่นักเรียนควรจะต้องตอบไว้ก่อนเป็นข้อ ๆ แล้วกำหนดคะแนนของแต่ละหัวข้อนั้น ๆ ตามลำดับความสำคัญมากน้อย เพื่อบังคับการให้คะแนนตามใจชอบ วิธีที่ง่าย ก็โดยเขียนคำตอบตัวอย่างสั้น ๆ ไว้ ว่านักเรียนจะต้องตอบให้ครอบคลุมแง่มุมใดบ้าง หรือจะเขียนเป็นตอน ๆ พร้อมกับคะแนนของแต่ละเนื้อความนั้น ๆ ก็ได้ ดังนี้

คำถาม : เหตุใดพันท้ายนรสิงห์จึงยินดียอมรับโทษประหารชีวิต ?

เพื่อจะรักษาความศักดิ์สิทธิ์ของกฎหมาย	10 คะแนน
เพื่อแสดงความจงรักภักดีต่อพระมหากษัตริย์	10 คะแนน
เพื่อแสดงว่าไม่ต้องการรับการยกเว้นจากกฎหมาย	10 คะแนน
เพื่อแสดงว่าทหารต้องไม่กลัวความตาย	8 คะแนน
เหตุผลอื่น ๆ ที่เหมาะสม	9 คะแนน
การรวบรวมเค้าโครงคำตอบ	 คะแนน
การร้อยกรอง	 คะแนน
อ้างเหตุผลผิด	หักคะแนนออกครึ่งละ 10 คะแนน

ข้อสอบแบบปรนัย

ข้อทดสอบแบบปรนัยเป็นเครื่องมือวัดผล ที่เพิ่งจะเกิดขึ้นเมื่อไม่นานมานี้เอง ต้นกำเนิดของข้อสอบชนิดนี้ ก็เพื่อประสงค์ที่จะแก้ไขข้อบกพร่องต่าง ๆ ของข้อสอบแบบปากเปล่าและแบบเรียงความ ซึ่งได้แก่การตรวจให้คะแนนไม่แน่นอนและการตอบแบบตบตาครู เป็นต้น ส่วนที่ว่าข้อสอบชนิดนี้จะสามารถปรับปรุงการสอบวัดได้จริงหรือไม่นั้น ก็ยังไม่เป็นที่รับรองกันทั่วไป ฝ่ายที่สนับสนุนข้อสอบแบบปรนัยก็ยืนยันว่าเขาสามารถพิสูจน์ได้ว่าข้อสอบแบบนี้มีความเที่ยงตรงและเชื่อมั่น สูงกว่าแบบเดิมมาก ส่วนฝ่ายถือหางแบบเรียงความก็อ้างว่า การสอบแบบปรนัยเป็นการส่งเสริมให้นักเรียนจดจำแต่เนื้อหาวิชา โดยไม่คำนึงถึงความมุ่งหมายอันแท้จริงของการศึกษา แต่ไม่ว่าจะเป็นเช่นไรก็ตาม มีการวิพากษ์วิจารณ์ถึงข้อสอบแบบปรนัยนั้น ส่วนใหญ่เกิดจากการเขียนข้อคำถามที่ผิดพลาด วางแผนการทดสอบบกพร่อง และการเลือกเนื้อหาที่ไม่เหมาะสมมาตั้งเป็นคำถามแทบทั้งสิ้น ถ้าตามที่กล่าวมานี้เป็นความจริง คำแนะนำและกฎเกณฑ์ง่าย ๆ ดังที่จะกล่าวต่อไปนี้ ก็จะสามารถช่วยครูให้ปรับปรุงข้อทดสอบแบบปรนัยให้ดียิ่งขึ้นได้มาก

ผู้อ่านอาจจะสังเกตว่า อนุสารนี้กล่าวขวัญถึงข้อสอบแบบปรนัยมากอยู่สักหน่อย ซึ่งทั้งนี้ก็โดยมีเหตุผลอยู่ 2 ประการคือ ประการแรก ครูทั่ว ๆ ไปไม่มีโอกาสได้ศึกษาแบบคำถามของข้อสอบชนิดนี้มากมายนัก และประการที่สองนั้น ก็เพราะมักจะยังไม่ทราบว่า การทดสอบแบบปรนัยนั้นมีขบวนการวิธีการสอบได้หลายทาง

ในบรรดาข้อสอบประเภทจำกัดคำตอบนั้น แบบที่นิยมใช้กันมากที่สุดก็คือแบบเลือกตอบ โครงสร้างทั่วไปของคำถามชนิดนี้ประกอบด้วย

สองส่วน คือ **ตอนนำ** ซึ่งเป็นตัวคำถามหรือเป็นข้อความที่จะให้นักเรียน
 กิจตอบ และ **ตอนตอบ** ซึ่งจะมีทั้งคำตอบที่ถูกปนกับอีก 3 หรือ 4
ตัวลวง หรือ **ตัวผิด** สำหรับครูที่เพิ่งเริ่มหัดเขียนคำถามแบบเลือกตอบ
 ใหม่ ๆ นั้น ข้อแนะนำให้เขียนตอนนำเป็นประโยคคำถามสมบูรณ์ ไม่ควร
 เขียนให้เป็นแบบอ่านต่อความเลย ทั้งนี้เพราะการเขียนเป็นคำถามสมบูรณ์
 มีโอกาสผิดพลาดน้อย เมื่อเขียนคำถามลงไปแล้วก็เริ่มเขียนตัวเลือก ซึ่ง
 จะมีจำนวนมากหรือน้อยขึ้นอยู่กับครูผู้เขียนเอง แต่โดยทั่วไป คำถาม
 ที่มีตัวเลือก 4 หรือ 5 ตัวจะดีกว่าชนิดที่มีเพียง 2 หรือ 3 ตัวแน่นอน เพราะ
 โอกาสที่จะ “เกา” ตัวคำตอบลงน้อยลง มีคำถามบางข้อเหมือนกันที่ครูไม่
 สามารถเขียนตัวเลือกที่เหมาะสม ๆ ได้มากกว่า 3 ตัว ถ้าเป็นเช่นนั้น ก็ควร
 เปลี่ยนไปใช้ปรนัยแบบอื่น ๆ เสีย โดยปรกติ ควรจะให้จำนวนตัวเลือกของ
 คำถามแบบเลือกตอบในข้อสอบแต่ละฉบับมีจำนวนเท่ากันเสมอ

ข้อแนะนำต่าง ๆ ดังที่จะกล่าวต่อไปนี้ มีความมุ่งหมายต่อคำถาม
 แบบเลือกตอบชนิดสามัญโดยเฉพาะ แต่ก็อาจนำไปใช้กับข้อสอบปรนัย
 แบบอื่น ๆ ได้เช่นกัน ซึ่งจะได้อธิบายโดยละเอียดในภายหลัง

(1) แบบฟอร์มของคำถามควรจะง่ายและสม่ำเสมอเป็นแบบเดียว
 กัน การเปลี่ยนแบบหรือชนิดของคำถามบ่อย ๆ ไม่เกิดประโยชน์อะไร
 เป็นพิเศษขึ้นมาเลย นอกจากจะทำให้เด็กต้องอ่านคำสั่งใหม่เท่านั้น
 มีคำถามหลายชนิดที่เราสามารถเปลี่ยนให้เป็นแบบเลือกตอบได้ทั้งนั้น แต่
 ทั้งนี้ก็ได้หมายความว่า เราจะไม่เปลี่ยนแบบของคำถามเสียเลยที่
 เดียว แต่หมายความว่า ควรจะเปลี่ยนแบบคำถามของข้อสอบแต่ละฉบับ

ข้อแนะนำเหล่านี้ดัดแปลงมาจากบทความในหนังสือ “Critical Thinking in Social
 Science”, แต่งโดย Paul Dressel และ Lewis Mayhew.

ให้น้อยที่สุด และควรเปลี่ยนก็ต่อเมื่อความมุ่งหมายในการสอบวัดตอนนั้นบังคับให้เราต้องเปลี่ยน ข้อสำคัญประการแรกของครูก็คือ **อย่าทำให้นักเรียนรู้สึกสับสน**

(2) เนื้อหาของคำถามที่อยู่ในตอนนำจะต้องแจ่มแจ้งจนนักเรียนสามารถทราบชัดว่าเขาจะต้องตอบอย่างไร เรื่องนี้จะต้องถือเป็นข้อปฏิบัติที่สำคัญ ไม่ว่าจะเขียนคำถามนั้นเป็นประโยคสมบูรณ์หรือเขียนเป็นแบบให้นักเรียนอ่านต่อความดังตัวอย่างนี้

ไม่ตี : ข้อความใดเป็นความจริง ?

ตีชั้น : ข้อความใดที่แสดงถึงสถานการณ์ต่อสู้ของมหาตมะคานธี ได้ถูกต้องที่สุด ?

(3) ตัวเลือกทุกตัวจะต้องใช้ภาษาหรือไวยากรณ์ ให้กลมกลืนกับตอนนำเสมอ เรื่องนี้ยังมีความสำคัญเป็นพิเศษ ถ้าตอนนำเขียนเป็นคำถามไม่สมบูรณ์ ถ้าจะว่ากันไปแล้ว คำถามที่มีตอนนำเป็นแบบประโยคไม่สมบูรณ์ หรือ ประโยคสมบูรณ์ก็ไม่มีคุณค่าอะไรยิ่งหย่อนกว่ากันเลย ฉะนั้นทางที่ดีที่สุดก็คือเลือกใช้แบบที่เหมาะสมกับเนื้อหาที่จะถาม โดยไม่ทำให้นักเรียนสับสนหรือใช้ถ้อยคำมากเกินไปจนจำเป็น

(4) คำถามจะต้องไม่ทำให้นักเรียนเข้าใจสับสนเพราะใช้คำฟุ่มเฟือย หรือใช้ประโยคที่ซับซ้อน เช่น ตัวอย่าง

ไม่ตี : ในตอนต้นศตวรรษที่ 19 เจ้าของโรงงานในอังกฤษได้ใช้วิธีการหลายอย่างที่จะเพิ่มผลผลิตด้วยราคาต่ำให้แก่ตัวเอง โดยกินแรงงานจากพวกกรรมกร ข้อความใดที่สนับสนุนคำดังกล่าวนี้

ตีชั้น : ข้อความใดที่แสดงว่ากรรมกรในประเทศอังกฤษในต้นศตวรรษที่ 19 ได้รับการบีบคั้น

(5) ภาษาที่ใช้ในข้อสอบจะต้องเหมาะกับระดับ การศึกษา ของ เด็กที่จะสอบ นั่นคือศัพท์ต่าง ๆ จะต้องไม่ยากเกินกว่าเด็กวัยนั้นจะฟัง ทราบ มิฉะนั้นแล้ว คำถามนั้นก็ท้อความยากโดยเพิ่มสิ่งอื่น ๆ เข้ามา อีก

(6) อย่าให้มีการแนะคำตอบ โดยอาศัยไวยากรณ์เป็นที่สังเกต (ในข้อสอบภาษาอังกฤษ) เรื่องนี้มักเกิดได้ง่ายในเมื่อตอนถามบ่งว่าคำตอบ ต้องเป็นคำพหูพจน์ และในตัวเลือกก็มีคำเพียงพยางค์เดียว หรือในเมื่อ ตอนนำลงท้ายด้วย a หรือ an เป็นต้น ทั้งตัวอย่างนี้

ไม่ดี : When students want general information about a subject they should consult an
 A. encyclopedia
 B. monograph
 C. dictionary
 D. lexicon
 E. newspaper

ดีขึ้น : When students want general information about a subject they should consult
 A. an encyclopedia
 B. a monograph
 C. a dictionary
 D. a lexicon
 E. an atlas

(7) ตำแหน่งของคำตอบถูกต้องกระจายจนนักเรียนไม่สามารถ จับได้ว่าตัวคำตอบจะต้องตกอยู่ที่ตัวเลือกสองตัวแรก หรือที่ตัวสุดท้าย หรือจะเวียนไปตามลำดับนั้นๆ. ก่อนที่จะรวบรวมข้อคำถามเข้าเป็นฉบับ จริงนั้น ครูควรจะสอบตัวคำตอบตกลงในเศษกระดาษก่อน เพื่อจะ ดูว่าเกิดการเวียนตัวตอบขึ้นหรือเปล่า (เช่น ข-ข-ข-ง-ง-ง-ก-ก-ก

เป็นต้น) และตัวถูกจะต้องกระจายไปตามตัวเลือกต่าง ๆ จนมีจำนวนกำลัง
กันหรือไม่อีกด้วย(คือจำนวนครั้งที่ตัวเลือก ก. ถูก จะต้องมีปานๆ กับตัว
เลือก ข. หรือ ค. เป็นต้น)

(8) ตัวเลือกทุกตัวจะต้องเป็นเอกพันธ์อยู่ในสกุลเดียวกัน และมี
ที่ทำพอที่จะเป็นตัวถูกได้ด้วย แต่ไม่ถึงกับถูกมากจนทำให้นักเรียนหลง
ผิดไปเลือกเป็นตัวตอบถูก.

(9) ใจความหรือเนื้อเรื่องที่จะถามนั้น จะต้องเป็นเรื่องราวที่ครู
สามารถเขียนตัวลงหรือตัวฝึกได้ครบ ถ้าปรากฏว่าเนื้อความตอนนั้น
มีที่ทำที่จะเขียนตัวลงได้เพียงตัวเดียวแล้ว ก็ไม่สมควรที่จะเขียนตัว
ลงซ้ำๆ กันเลย เพราะถึงจะสร้างคำถามซ้อนนั้นได้สำเร็จ นักเรียนก็จะ
เลือกตัวถูกซึ่งแปลกกว่าตัวลงอื่น ๆ ออกมาได้โดยง่ายดังเช่น

ไม่ตี : ท่านจะพบสิ่งใดในวัดไทย ?

- ก. โบสถ์
- ข. โคม
- ค. สุเหร่า
- ง. ศาลพระภูมิ
- จ. ไม้กางเขน

ตีขึ้น : ท่านจะไม่พบอะไรอย่างหนึ่งในวัดไทย จงเลือกสิ่งนั้นตอบ.

- ก. โบสถ์
- ข. เจดีย์
- ค. หอไตร
- ง. สุเหร่า
- จ. ศาลา

(10) ทั้งตัวเลือกที่ถูกและผิดจะต้องเขียนด้วยภาษาที่มีความชัดเจนปานๆ กันคล้ายๆ กัน ถ้าตัวคำตอบจำต้องขยายด้วยวลีต่างๆ จนมีข้อความยืดยาวแล้ว ตัวลวงก็ต้องขยายให้ยาวตามนั้นด้วย ถ้าตัวลวงเขียนเพียงสั้นๆ โดยไม่ระมัดระวังแล้ว นักเรียนก็จะสังเกตได้ทันทีว่าตัวเลือกที่เขียนไว้อย่างรัดกุมนั้น มักจะต้องเป็นตัวเลือกเสมอ จึงจำเป็นที่ครูผู้สอนจะต้องเขียนตัวลวงให้มีความยาวใกล้เคียงกับตัวเลือกเสมอ

(11) คำถามแต่ละข้อ ควรจะถามแต่ใจความที่สำคัญๆ เท่านั้น สำหรับข้อที่ครูตั้งใจจะให้เป็นอย่างนั้น จะต้องไม่ใช่ยากเพราะถามคลุมเครือหรือถามในสิ่งปลีกย่อย แต่เป็นข้อยากเพราะนักเรียนต้องนำความรู้ด้านต่าง ๆ มาเปรียบเทียบกัน แล้วนำไปพิจารณากับตัวเลือกที่เขียนไว้แล้วเป็นอย่างดีเท่านั้น

(12) จงเว้นการใช้คำบางคำ ที่จะแนะนำให้เด็กตัดทิ้งตัวลวงออก เช่นคำว่า เสมอ เท่านั้น ทุก ๆ ฯลฯ แต่อย่างไรก็ดี บางที คำเหล่านั้นก็มีประโยชน์สำหรับถกกันนักเรียนที่ “กล่อกการสอบ” ซึ่งมักจะเลือกตัวเลือกที่มีคำเหล่านั้น เพราะรู้ยู่ว่าเป็นลวงที่เด่นชัดเกินไป เลยทำให้เขาไปตอบตัวผิดอื่นๆ

(13) ไม่ควรใช้คำถามที่มีคำปฏิเสธซ้อน เพราะรังแต่จะทำให้เด็กคิดสับสนเปล่าๆ เช่นถ้ามีคำว่า “ไม่” หรือ “ผิด” อยู่ในตอนนำแล้ว ก็จงอย่าใช้คำปฏิเสธอื่นๆ เพิ่มในตอนนำหรือในตัวเลือกอีกเลย วิธีที่ดี ก็โดยย้ายถามแบบปฏิเสธเหล่านั้นมาอยู่เสียด้วยกัน หรืออาจตัดคำปฏิเสธนั้นออกเสียเลย โดยกล่าวถามซ้ำอีกทีหนึ่งก็ได้ ดังเช่นคำถามนี้ :

ข้อความต่อไปนี้เกี่ยวกับเรื่องการเขียนภาพในสมัยเรณอของ ซึ่งกล่าวถูกต้องทั้งสี่ แต่มีผิดอยู่ข้อเดียวจงหาข้อที่ผิดนั้น. หรือข้อความนี้กล่าวถึงการเขียนภาพในสมัยเรณอของถูกต้องหมดทุกข้อ นอกจาก.....

(14) ข้อความประเภท “ผิดหมดทุกข้อ” หรือ “ที่กล่าวมาผิดหมด” เหล่านี้ ก็พอจะใช้เป็นตัวเลือกได้เหมือนกัน และมีประโยชน์ที่จะเตือนให้นักเรียนตรวจตราตัวเลือกอื่นๆ ให้รอบคอบยิ่งขึ้น แต่ในแบบทดสอบที่ใช้ตัวเลือกชนิดนี้ ก็จะต้องให้ตัวเลือกนั้นมีโอกาสถูกหรือผิดด้วยเช่นกัน

(15) ตัวเลือกประเภท “ถูกหมดทุกข้อ” หรือ “ที่กล่าวมาถูกหมด” ก็อาจใช้เป็นตัวเลือกได้เช่นกัน แต่มีข้อบกพร่องตรงที่มักจะกลายเป็นการแนะนำคำตอบไปเสีย ถ้านักเรียนเห็นว่าเพียงแต่ตัวเลือก 2 ตัวแรกก็เป็นจริงเสียแล้ว ฉะนั้นถ้าผสมทั้ง “ถูกหมดทุกข้อ” กับ “ผิดหมดทุกข้อ” ไว้ภายในข้อเดียวกันก็จะมีคุณค่ามากขึ้น. เมื่อนำตัวเลือก “ถูกหมดทุกข้อ” ไปใช้ก็ควรระวังตัวไว้ 2 ประการคือ ถ้าใช้ตัวเลือกนี้ปนกัน “ผิดหมดทุกข้อ” แล้ว จะต้องให้ตัวเลือก “ถูก.....” อยู่หน้า “ผิด.....” ก่อนเสมอ และอีกประการหนึ่งนั้น จะต้องกล่าวเน้นไว้ในคำชี้แจงว่า ให้นักเรียนเลือกคำตอบที่ **ดีที่สุด** หรือ **ถูกต้องที่สุด** ลงไว้ด้วยเสมอ ถ้ามีฉะนั้นแล้ว นักเรียนก็อาจแย้งได้ว่าตัวเลือกตัวอื่น ๆ ก็เป็นคำตอบถูกต้องด้วยเหมือนกัน

(16) ข้อคำถามทั้งหมดควรจะได้สอบทาน โดยเพื่อนครูสักคนหนึ่งหรือหลาย ๆ คน หรือกับครูผู้ช่วยก็ได้ เพื่อให้แน่ใจ ว่าคำถามเหล่านั้นชัดเจน เหมาะสมและถูกต้องแล้ว.

(17) ควรแจกแจงข้อคำถามให้ตามวัตถุประสงค์หรือทักษะต่างๆ จนครบ และให้มีจำนวนข้อกับความมุ่งหมายของการสอบครั้งนั้น ๆ ด้วย

(18) ตามปรกตินั้น คำถามของข้อสอบแต่ละฉบับจะต้องมีความยากง่ายต่าง ๆ กันตั้งแต่ก่อนข้างง่ายจนถึงก่อนข้างยาก โดยให้มีข้อที่ยากปานกลางมาก ๆ ข้อ

(19) ควรตรวจราคาคำถามในแบบทดสอบให้ถี่ถ้วนดี จนมั่นใจว่า ข้อความที่กล่าวไว้ในข้อหนึ่งข้อใดจะต้องไม่ไปแนะคำตอบของข้ออื่นๆเลย

(20) ควรพิจารณาให้แบบทดสอบมีความยาวพอเหมาะ จนเด็กส่วนใหญ่สามารถทำเสร็จภายในเวลาที่กำหนดให้

(21) ต้องตรวจตราแบบทดสอบ ในเรื่องเครื่องหมายวรรคตอน และความคลาดเคลื่อนในไวยากรณ์และการพิมพ์ด้วย. การพิมพ์ข้อสอบ สำคัญกว่าการพิมพ์แบบเรียนเป็นอันมาก เพราะความบกพร่องนี้จะทำให้ ความหมายเพี้ยนผิดแผกไปจากความหมายที่ตั้งใจจะถาม

(22) หลังจากการสอบแล้ว ควรนำคำถามเหล่านั้นมาปรับปรุง ให้ดีขึ้นโดยยึดเอาคำตอบของนักเรียนเป็นหลักสำหรับพิจารณา การนำ ข้อสอบมาอภิปรายกับเด็กนักเรียนในชั้น สามารถช่วยให้การปรับปรุง ข้อสอบและการสอบในอนาคตดีขึ้น.

ก่อนที่จะลงมือเขียนข้อทดสอบชนิดนี้ ก็มีวิธีปฏิบัติเบื้องต้นที่สำคัญๆ อยู่หลายขั้น ทำนองเดียวกันกับการออกข้อสอบประเภทเรียงความ หรือปากเปล่า. ครูไม่ควรจะออกข้อสอบโดยวิธีตั้งหน้าตั้งตาเขียน คำถาม เรื่อยไป จนเห็นว่าได้มากข้อพอกับเวลาสอบเช่นนี้เลย แต่ครูควรจะเริ่ม ด้วยทำตาราง ดังที่แสดงไว้ในหน้า 20 เพื่อจะได้มั่นใจว่าข้อสอบฉบับ นี้ครอบคลุมทั้งเนื้อหาและความมุ่งหมายของวิชานั้นเป็นสมมูลย์กัน และ ในการนี้เรามักจะพบอยู่เสมอๆ ว่า มีคำถามหลายข้อที่เกี่ยวที่เราอาจจะ จัดให้เป็นคำถามที่วัดความมุ่งหมายชนิดนั้นหรือชนิดนั้นก็ได้ ฉะนั้นเราจึง ควรถือว่า ตารางข้อสอบมีหน้าที่เพียงบอกเค้าโครงทั่วไปเท่านั้น

ในการรวบรวมคำถามต่าง ๆ เข้าเป็นฉบับข้อสอบนั้น ก็ควรถือ หลักปฏิบัติดังนี้ ถ้าข้อสอบฉบับนั้นมีคำถามหลายชนิด (ถูก-ผิด, เลือก

ตอบ ฯลฯ) ก็ควรจะรวมแต่ละชนิด ๆ ไว้ด้วยกัน เพื่อมิให้นักเรียนสับสน และควรจะกำหนดให้เป็นอย่างเดียวกันด้วย ว่าจะให้นักเรียนเลือกตอบอย่างไร เช่นจะให้นักเรียนเลือกเพียงคำตอบที่คิดที่สุดเพียงตัวเดียวตามที่ปฏิบัติกันทั่วไป หรือจะให้ตอบข้อละหลาย ๆ ตัวเป็นต้น สำหรับวิธีตอบแบบหลังนี้มีความยุ่งยากอยู่สักหน่อย ตรงที่ตัวผิดจะต้องเขียนให้ผิดอย่างเด่นชัดจริงๆ ส่วนตัวถูกเหล่านั้นจะต้องเขียนให้มีปริมาณความถูกมากถูกน้อยลดหลั่นไม่เท่ากัน. อนึ่งเพื่อผลทางจิตวิทยา เขานิยมเริ่มต้นคำถามด้วยข้อง่าย ๆ แล้วค่อย ๆ ทวีความยากขึ้นไปตามลำดับจนจบแบบทดสอบฉบับนั้น.

เนื่องจากภาษาที่ใช้ในข้อสอบปรนัยก็มีความสำคัญมาก เขาจึงแนะนำให้รวบรวมข้อสอบไว้ล่วงหน้าหลาย ๆ วันก่อนการสอบจริง วิธีนี้จะช่วยให้ครูสามารถตรวจทานคำถามเหล่านั้น (หรือจะให้เพื่อนร่วมงานตรวจด้วยก็จะยิ่งดี) เพื่อขจัดคำถามที่เป็นปัญหาสองแง่ คำถามที่ไม่เหมาะสม และตรวจตราความถูกต้องของคำถามเป็นต้น.

ข้อสอบแบบถูก-ผิด

ข้อสอบชนิดนี้ครูใช้กันแพร่หลายมาก เพราะสร้างได้ง่าย และตรวจให้คะแนนง่าย ข้อสอบแบบ ถูก - ผิด มีความบกพร่องที่สำคัญประการแรกตรงที่ แต่ละข้อมีโอกาสให้เลือกเพียงสองทาง ซึ่งไม่สามารถป้องกันการได้คะแนนสูงเนื่องจากการเดาได้. ข้อบกพร่องประการที่สองที่ผู้ออกข้อสอบประสบก็คือจะต้องหาข้อความที่ไม่สมบูรณ์เกินไป คือไม่จริงหรือเท็จจนเด่นชัดเกินไป — คือถูกหรือผิดจนแจ้งชัด แต่การที่จะรู้ว่าสิ่งใดถูกหรือผิด แม่นยำหรือบกพร่อง เหล่านี้ก็เป็นความมุ่งหมาย

ที่สำคัญประการหนึ่งของการศึกษา ฉะนั้นสิ่งที่ควรจะทำในข้อสอบชนิดนี้ก็คือ จะต้องหาวิธีใดวิธีหนึ่งไปเพิ่มสมรรถภาพข้อสอบแบบถูก—ผิด ให้ใช้การได้ดียิ่ง ๆ ขึ้น ดังเช่น อาจจะทำให้โดยวิธีใช้ตัวเลือกกที่ กั้น ตัวอย่าง คำถามจากข้อ 1 — 2 เกี่ยวกับคำสั่งสอนของพระพุทธเจ้าแต่ละข้อประกอบไปด้วย ข้อความ และ เหตุผล. จงเลือกจากบัญชี ก.—จ. ข้างล่าง ว่าข้อความและเหตุผลแต่ละคู่กันสัมพันธ์กันอย่างไร ก็ตอบดังนี้

- ก. ทั้งข้อความและเหตุผล ถูกต้อง แต่ไม่เป็นสาเหตุและผลต่อกัน
- ข. ทั้งข้อความและเหตุผล ถูกต้อง และเป็นเหตุผลต่อกัน
- ก. เฉพาะข้อความ ถูกต้อง แต่เหตุผล ผิด (ไม่สนับสนุนกัน)
- ง. เฉพาะข้อความ ผิด แต่เหตุผล ถูก (ไม่สนับสนุนกัน)
- จ. ทั้งข้อความและเหตุผล ผิด

วิธีทำ ครั้นแรกจงคิดดูก่อนว่าข้อความนั้นถูกต้องหรือผิด — แล้วคิดว่าเหตุผลนั้นถูกหรือผิด — จากนั้นจึงตอบว่าอยู่ในประเภทใดจาก ก. — จ.

(1) ข้อความ สัจธรรมมีอยู่จริง

เหตุผล เรื่องนี้อาจพิสูจน์ได้โดยประสาธน์สัมผัส

(2) ข้อความ นักปราชญ์มักกลัวตาย

เหตุผล เพราะเขาทราบว่าเขตายแล้วจะต้องสูญ

จะเห็นว่า คำถามแบบนี้ซับซ้อนเกินไปสำหรับเด็กชั้นประถม ควรจะใช้กับนักเรียนมัธยมต้นหรือมัธยมปลายจึงจะเหมาะ ยังมีวิธีแปลงคำถามแบบถูก—ผิดแบบดั้งเดิมนี้ให้สามารถวัดความรู้และผลการเข้าใจอีกหลายแบบ ดังแบบที่ให้ให้นักเรียนตรวจแก้คำผิดในข้อความนั้น เช่น

ตัวอย่าง Shale เป็นหินตะกอนที่มี calcite มากที่สุดคำกล่าวนี้เป็นเช่นไร?

- ก. ถูกต้องแล้ว
- ข. ยังไม่ถูกต้อง แต่ถ้าเปลี่ยนคำ "Shale" เป็น "Slate" แล้วจึงจะถูกต้อง
- ค. ยังไม่ถูกต้อง แต่ถ้าเปลี่ยนคำ "หินตะกอน" เป็น "หินอัคนี" แล้วจึงจะถูกต้อง
- ง. ยังไม่ถูกต้อง แต่ถ้าเปลี่ยนคำ "calcite" เป็น "clay" แล้วจึงจะถูกต้อง

จ. ยังไม่ถูกต้อง และการเปลี่ยนทั้งกล่าวก็ไม่มีข้อใดถูกเลย

คำถามแบบนี้ มีคุณค่าเพิ่มตรงที่นักเรียนจะต้องเลือกคำตอบจาก 5 ทางซึ่งจะทำให้โอกาสเผลอถูกลดน้อยลงมาก

วิธีพลิกแพลงอีกแบบหนึ่ง ก็โดยให้นักเรียนตรวจสอบความถูกต้องก่อนว่าถูกหรือผิด แล้วให้นักเรียนชี้แจงด้วยว่า **ทำไม** ข้อความนั้นจึงเป็นเช่นนั้น ในคำถามแบบนี้ เราอาจลดความสับสนของเด็กลงได้มาก ถ้าขีดเส้นใต้คำที่สำคัญไว้ให้เด่นชัด อย่างไรก็ตามก็ถ้าครูยังประสงค์จะใช้คำถามแบบ ถูก—ผิด ชนิดธรรมดาอยู่อีก ก็มีหลักบางประการที่ควรจะช่วยทำให้เครื่องมือวัดชนิดนี้มีความเชื่อมั่นได้ดียิ่งขึ้น คือ

(1) จงระวังการใช้คำพูดให้จงหนัก เพราะอาจทำให้คำถามแบบ ถูก—ผิด กลายเป็นคำถามที่มี "ลูกไม้" ได้ง่ายๆ ถ้าเป็นคำถามที่มีข้อความมากกว่าหนึ่งใจความ จะต้องระวังให้ทั้งสองใจความ (หรือทุกๆ ใจความ) นั้น ถูกหมดหรือผิดหมดด้วยเสมอ และเพื่อจะป้องกันมิให้นักเรียนสับสน เขาจึงแนะนำให้เขียนคำถามเพียงข้อละใจความเท่านั้นเพื่อจะได้แน่ใจ

ว่าเป็นการวัดความรู้จริง ๆ มิใช่เป็นการวัดทักษะในการอ่านไปเสีย ดัง
ตัวอย่างนี้

ไม้ดี รัฐธรรมนูญฉบับใหม่ของประเทศไทย กำหนดให้มีสภานิติบัญญัติ
2 สภา ซึ่งในรัฐธรรมนูญฉบับที่แล้วมิได้มีบทบัญญัติเกี่ยวกับสภา
นิติบัญญัติไว้โดยเฉพาะ

ศิรินทร์ รัฐธรรมนูญฉบับใหม่ของไทยกำหนดให้มีสภานิติบัญญัติ 2 สภา

(2) มีคำบางคำหรือบางวลีที่แนะนำคำตอบให้กับเด็ก ดังเช่น คำว่า
“ไม่เลย, ทุกครั้ง, มิใช่, ทั้งสิ้น ฯลฯ” ถ้าเราใช้คำเหล่านี้ในคำถาม
ก็มักจะเป็นการแนะนำนักเรียนว่า ข้อความนั้นผิด ส่วนคำถามที่ใช้คำจำพวก
“โดยทั่วไป—ตามธรรมดา—ตามปกติ ฯลฯ” เหล่านี้ ก็มักแนะนำว่าข้อความ
นั้นถูก ฉะนั้นถ้าจะต้องใช้คำเหล่านี้แล้ว ก็ควรใช้ในลักษณะที่จะลง
ให้เด็กที่ชอบเถียงคำตอบโดยสังเกตคำเหล่านี้ตอบผิด

ตัวอย่าง

ไม้ดี กฎหมาย กำหนดให้พระมหากษัตริย์ ของไทยเพียงพระองค์เดียว
เท่านั้นที่ต้องเป็นพุทธมามก

ศิรินทร์ รัฐธรรมนูญของประเทศไทยกำหนดให้ในหลวง ทุกพระองค์ ทรง
เป็นพุทธมามก

(3) ผู้เขียนคำถามชนิดนี้ควรจะต้องระมัดระวังถ้อยคำหรือวลีที่
มีความหมาย 2 นัยจนหนัก เพื่อป้องกันมิให้นักเรียนหลงเข้าใจผิด เพราะ
ถ้าเด็กต้องมัวแต่ไปนึกตีความหมายของคำบางคำเสียแล้ว ความสามารถ
ที่จะนำความรู้มาใช้ในการตอบก็จะค่อยลง ความบกพร่องในเรื่องนี้ก็จะ
จะได้แก่การใช้คำคุณศัพท์และวิเศษณ์มากเกินไป เช่นใช้คำว่า บ่อยๆ—

น้อยครั้ง—นาน ๆ ครั้ง เป็นต้น เด็กแต่ละคนจะตีความหมายของคำเหล่านี้ไม่เหมือนกัน เพราะไม่เข้าใจว่า จะต้องมากสักเพียงใดจึงจะเรียกว่าบ่อย ๆ เป็นต้น ดังตัวอย่างนี้

ไม่คิด เราไม่ค่อยจะได้พบนกที่ชั้นเพราะ ๆ

คิดขึ้น แถบข้าวโลกมีนกที่ชั้นเพราะ ๆ น้อยกว่าแถบศูนย์สูตร

ข้อสอบแบบจับคู่หรือจัดเข้าพวก

ข้อสอบปรนัยชนิดสามัญแบบที่สองก็คือแบบที่ให้ **จับคู่ และ จัดเข้าพวก** ข้อสอบชนิดนี้เหมาะสำหรับถามเกี่ยวกับ ชื่อ วัน เวลา คำศัพท์ ความคิดเห็น เป็นต้น ตามจริงนั้น เราอาจนำวิธีจับคู่ไปใช้กับคำถามที่เป็นภาพ กราฟ แผนที่ รูปภาพ หรือแผนผังต่าง ๆ ก็ได้. แต่โดยทั่วไปมักจะเป็นแบบให้จับคู่ระหว่างข้อความในแถวแรกกับในแถวที่สอง ความบกพร่องที่พบเสมอ ๆ ในข้อสอบแบบนี้ ก็ได้แก่การนำเอาเรื่องราวต่าง ๆ ที่เป็นวิวิธพันธุ์มาก ๆ เรื่องมาปนกันจนยาว โดยนึกว่าจะเป็นเรื่องบ๊องกันหรือลดการเคาของนักเรียนให้น้อยลง ดังตัวอย่างในหน้า 40

ผู้ใช้คำถามชนิดนี้จะต้องระวังให้เป็นคำถามที่มุ่งวัดความรู้ของนักเรียนโดยตรง มิใช่วัดความสามารถในการตอบถูกด้วยวิธีตัดข้อความที่ไม่เกี่ยวข้องออกไปทีละตัว ๆ จนพบคำตอบ ดังเช่นในข้อ 1 แถว 2 นั้น นักเรียนจะจับตรงคำ “แต่งตั้งใหม่” ได้ ว่าจะต้องเป็นข่าวที่ปรากฏในหน้าหนังสือพิมพ์เร็ว ๆ นี้เอง และยิ่งได้รับคำแนะนำ “สทรรฐ” มาประกอบด้วยแล้ว ก็จะทำให้เหลือตัวเลือกที่น่าจะเป็นคำตอบถูกเพียง 2 ตัวเท่านั้น เป็นการไม่ได้นำความรู้มาตอบเลย. ในข้อ 2 ก็มีการแนะนำคำตอบแบบเดียวกันอีกที่คำว่า “ผู้นำฝรั่งเศส” และ “เทอโกลด์” ซึ่งเป็นชื่อชาวฝรั่งเศส และมีอยู่เพียงข้อเดียวในแถว 1. ส่วนข้อ 3 นั้น นักเรียนอาจหาคำตอบถูกโดยวิธีไล่ลงมาทีละตัว ๆ จนพบ

แถว I

- ก. ถนอม กิตติขจร
- ข. ไทวท์ ไอเซ็นเฮาว์
- ค. เอริโอเปีย
- ง. ฟิเดลคาสโตร
- จ. สำนักวาติกัน
- ฉ. ชาลส์ เคอโกล์
- ช. อาเธอร์ โกลด์คเบอก
- ซ. ประภาส จารุเสถียร
- ณ. เกษตรศาสตร์
- ญ. มอสโคว
- ฎ. ซีโต้
- ฏ. กาลอส รอมัลลอส
- ฐ. ยิบรอลต้า
- ท. นิวยอร์ก
- ฅ. กลองสวเช

แถว II

- 1. ผู้แทนสหรัฐประจำองค์การสหประชาชาติ ซึ่งได้รับการแต่งตั้งใหม่เมื่อเร็ว ๆ นี้
- 2. ผู้นำฝรั่งเศสซึ่งขู่ว่าจะลาออกจากองค์การซีโต้
- 3. นักปฏิวัติผู้นำระบอบคอมมิวนิสต์มาใช้ในประเทศของตน

เราอาจปรับปรุงคำถามแบบจับคู่ให้มีประสิทธิภาพยิ่งขึ้นถ้าจะจำกง่าย ๆ เหล่านี้ไว้ คือ

- 1. ควรจะกำหนดให้คำถามแต่ละชุด ประกอบด้วยเรื่องราวที่เป็นเอกพันธ์ หรืออยู่ในสกุลเดียวกันเพียง 5 ประการในแถวแรก และอีกแถวให้มีเพียง 3, 4, หรือ 6 ตัว. อย่าให้คำถามมีจำนวนเท่ากับตัวคู่เป็นอันขาด เพราะเด็กจะตอบข้อสุดท้ายถูกต้องโดยไม่ต้องคิดเลย

2. ต้องชี้แจงให้นักเรียนทราบก่อนว่า เขาจะต้องใช้ตัวที่จับคู่เป็นคำตอบได้เพียงครั้งเดียวหรือหลายครั้ง ถ้าให้ใช้ได้หลายครั้ง ก็อาจให้ตัวคู่มีจำนวนเท่ากับตัวถามได้

3. ควรพยายามตกแต่งถ้อยคำให้เข้าใจง่ายที่สุดเพื่อให้นักเด็กสับสนดังตัวอย่างคำถามต่อไปนี้

แถว 1

ก. สมัยสุโขทัย

ข. สมัยกรุงศรีอยุธยา

ค. สมัยกรุงธนบุรี

ง. สมัยกรุงรัตนโกสินทร์ก่อนการ

เปลี่ยนแปลงการปกครอง

จ. สมัยกรุงรัตนโกสินทร์หลังการเปลี่ยน

แปลงการปกครองจนถึงปัจจุบัน

แถว 2

1. ก้นพระพุทธรูปพระพุทธรูป จันทบุรี

สระบุรี

2. การเลิกทาส

3. การติดต่อกับประเทศจีนเป็นครั้งแรก

แรก

ข้อสอบปรนัยที่คล้ายกับแบบจับคู่อีกชนิดหนึ่งก็คือ “แบบตัวเลือกคงที่” หรือ “แบบจัดเข้าพวก” ตัวเลือกที่กำหนดให้เป็นตัวคงที่นั้น อาจเป็นคาบเวลา, ความคิดรวบยอด, คำศัพท์, เหตุการณ์, บุคคล ฯลฯ ก็ได้ แล้วตั้งคำถามที่เกี่ยวกับสิ่งที่กำหนดให้ไว้นั้น ดังตัวอย่างนี้

ตัวอย่าง จงจับคู่กิจการงานข้างล่างกับสถานที่ราชการที่กำหนดไว้จาก ก-จ แล้วตอบดังนี้

ก. เป็นหน้าที่ของอำเภอโดยเฉพาะ

ข. เป็นหน้าที่ของจังหวัดโดยเฉพาะ

ค. เป็นหน้าที่ของอำเภอหรือจังหวัดก็ได้

ง. เป็นหน้าที่ของอำเภอและจังหวัดร่วมกัน

จ. ไม่ใช่หน้าที่ของอำเภอหรือจังหวัด ทั้งสองประการ

คำถาม

1. การสมรส
2. การตัดถนน
3. นโยบายต่างประเทศ

ฯลฯ

ยังมีวิธีตามอีกแบบหนึ่งสำหรับวัดความสามารถในการที่จะนำเอาเหตุการณ์, บุคคล, ฯลฯ ไปสัมพันธ์กับยุคสมัยทางประวัติศาสตร์โดยใช้เส้นตรงแทนระยะเวลา ดังนี้

ตัวอย่าง

บนเส้นข้างล่างนี้ ตัวอักษร ก.—จ. แทนคาบระยะเวลาของยุคต่าง ๆ ในประวัติศาสตร์ดังที่ได้เขียนไว้ใต้เส้น จงตอบว่าเหตุการณ์ในคำตอบข้อ 1—5 เกิดขึ้นในระยะใดตรงกับอักษรใด

ก : ข : ก : ง : จ

สงครามปฏิวัติ สงครามกลางเมือง สงครามเสปนญ—อเมริกา สงครามโลกครั้งที่ 1

คำถาม

1. การปฏิวัติฝรั่งเศส
2. รัชกาลที่ 5 เสด็จราชย์
3. เซ็กสเปียร์เขียนโอเทลโล
4. มุขโสลินีเป็นผู้เผด็จการฟาสซิสต์
5. มาร์คซ์เขียนตำราประกาศลัทธิคอมมิวนิสต์

เราอาจใช้หลักนี้ไปสร้างคำถามสำหรับวัดความรู้นักเรียนในด้านความสัมพันธ์ของเวลา. ยุคสมัย หรือการลำดับกาลเวลาก็ได้ เช่น
ตัวอย่าง

จงตอบคำถามข้อที่ 1—3 โดยขีดตอบดังนี้

- ก—ถ้าเป็นเหตุการณ์ที่เกิดก่อน พ.ศ. 1500
- ข—ถ้าเป็นเหตุการณ์ที่เกิดระหว่าง พ.ศ. 1500—1800
- ค—ถ้าเป็นเหตุการณ์ที่เกิดระหว่าง พ.ศ. 1800—2200
- ง—ถ้าเป็นเหตุการณ์ที่เกิดระหว่าง พ.ศ. 2200—2500
- จ—ถ้าเป็นเหตุการณ์ที่เกิดขึ้นหลัง พ.ศ. 2500

คำถาม

1. การสละราชสมบัติของพระปกเกล้า
2. นโปเลียนโบนาปาร์ตเป็นจักรพรรดิฝรั่งเศส
3. วุคโรวินสันได้รับเลือกเป็นประธานาธิบดีครั้งแรก

เวลา

ในการวัดความรู้ของนักเรียนเกี่ยวกับกาลเวลา เราอาจจัด
แปลงคำถามให้เป็นอีกแบบหนึ่ง ดังนี้

- ก. พระมงกุฎเกล้า
- ข. พระนั่งเกล้า
- ค. พระเจ้าอยู่หัวภูมิพล
- ง. พระพุทธเลิศหล้านภาลัย
- จ. พระจุลจอมเกล้า

แล้ววางคำตอบให้นักเรียนเลือกว่าจะต้องเรียงกันอย่างไรจึงจะถูกต้อง

1. ก ข ค ง จ

2. ง ข จ ก ก

3. ก ง จ ข ก

4. ข ก ง จ ก

5. จ ง ก ค ข

แบบนี้ก็ยังไม่ดีมากนัก เพราะเด็กยังอาจตอบถูกด้วยวิธีตัดตัวผิดออกไปทีละตัว ๆ ก็ได้ ถ้าเขามีความรู้เกี่ยวกับชื่อของคนแรกและคนหลังจริง ๆ เพียง 2 คนเท่านั้น

การใช้กราฟและรูปภาพประกอบคำถาม

จงใช้ภาพในข้อ 1–5 ตอบว่า ภาพใดเป็นกราฟที่แสดงความสัมพันธ์ชนิดใด?

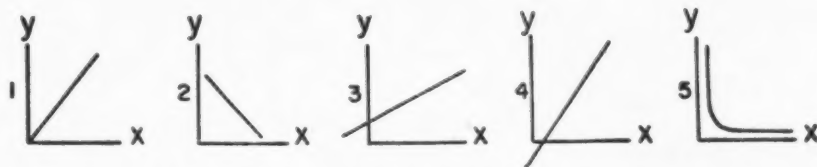
ก. สัมพันธ์อย่างโดยตรงประการเดียว

ข. เป็นสัดส่วนอย่างตรง

ค. สัมพันธ์อย่างกลับกันประการเดียว

ง. เป็นสัดส่วนอย่างกลับ

จ. ไม่มีข้อใดถูก



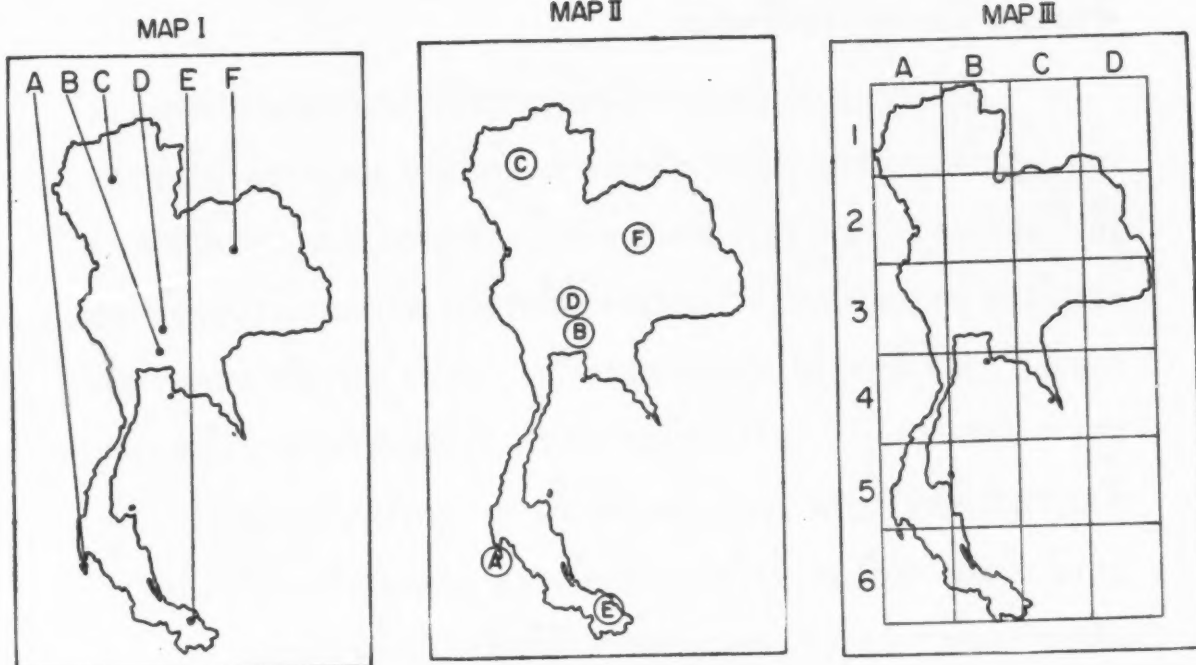
ถ้าจะทดสอบความรู้วิชาภูมิศาสตร์นั้น เราก็อาจตั้งคำถามให้ไปสัมพันธ์กับแผนที่ได้หลายแบบ ดังในภาพที่ 1 เป็นแบบที่เราใช้ลูกศรหรือเส้น ชี้ที่ตำบลต่าง ๆ เอาไว้ แล้วครุถามให้นักเรียนตอบถึงที่ตั้งของสถานที่ต่าง ๆ เช่น

ลูกศรใดแสดงที่ตั้งของมหาวิทยาลัยขอนแก่น

หรือจะถามถึงตำบลที่เกิดของเหตุการณ์ที่สำคัญๆ หรือภูมิภาคที่มีความสำคัญอะไรๆ ก็ได้ เช่น พระนเรศวรทรงทำยุทธหัตถีครั้งสำคัญตรงบริเวณลพบุรี ก-ข-ค-ง-จ เป็นต้น หรือเราจะใช้แผนที่ดังแบบที่ 2 ถามคำถามแบบเดียวกันนั้นก็ได้ บางทีครูอาจจะใช้รูปภาพแผนที่ที่มีการวางผังในแบบที่ 3 ถามให้นักเรียนชี้ตอบบริเวณหรืออาณาเขตนั้นๆ แทนที่จะให้ชี้จำเพาะเจาะจงลงไปเช่น

1. นักเรียนจะพบเผ่าชาวเขาในภูมิภาคแถบใด ?

1. ก
2. ข
3. ค
4. ง
5. ไม่มีข้อใดถูก



หรือจะถามว่า ในภูมิภาค 3 ข. มีผลิตผลอะไรที่สำคัญ ?

ก. ข้าว ข. ไม้สัก ค. ติบูก ง. ผลไม้ จ. ถั่วลิสง

อย่างไรก็ดี เราอาจตั้งคำถามวิชาภูมิศาสตร์โดยไม่ต้องใช้แผนที่ประกอบด้วยเลยก็ได้ ดังตัวอย่าง

พระนเรศวร กล่าวข้อความนี้กับพระมหาอุปราช ฦ ที่ใด
“เจ้าพี่จะมัว-เฉยอยู่ทำไม จงออกมาชนช้างกันให้ปรากฏเป็นเกียรติประวัติ
สืบไปเถิด”

- ก. ตำบล ตำนเจดีย์
- ข. ตำบล หนองสาหร่าย
- ค. ตำบล เขาชนไก่
- ง. ตำบล โพธิ์สามต้น
- จ. ตำบล ลาดหญ้า

คำถามที่ใช้บทความเป็นหลัก

มีข้อสอบปรนัยอีกแบบหนึ่งที่สามารถใช้วัดทักษะในการอ่านของเด็ก รวมทั้งความสามารถในการวิเคราะห์, ตัดสินใจ, และการนำไปใช้ ฯลฯ ได้เป็นอย่างดี ในการสร้างคำถามชนิดนี้ ครูต้องไปเลือกบทความบางตอนที่เหมาะๆ และตรงกับความประสงค์ที่จะวัดมาสักตอนหนึ่ง เช่น ถ้าต้องการวัดความเร็วในการอ่านและความเข้าใจ เราก็ต้องไปเลือกบทความที่ยาว ๆ สักหน่อย หรือถ้าต้องการจะวัดความสามารถในการวิเคราะห์ สังเคราะห์ ฯลฯ บทความนั้นก็ควรจะสั้น ๆ และไม่ยาวจนเกินไปจนเกินไป ครูจะต้องการวัดอย่างไรก็ตาม จะต้องคำนึงถึงเวลาให้เพียงพอ เพื่อนักเรียนจะได้มีโอกาสแสดงความรู้ความสามารถของเขาออกมาได้อย่างเต็มที่ด้วยเสมอ

เราอาจพลิกแพลงคำถามชนิดนี้ได้หลายประการ ทั้งจะยกตัวอย่างให้เห็นบ้างดังนี้

ทดสอบความสามารถในการรวบรวมหรือจัดลำดับ : กรุณากำหนดข้อความให้สัก 3 หรือ 4 ตอน โดยมีอักษร ก-ข-ค-ง-จ กำกับแต่ละตอนไว้แล้วถามว่า

ควรจะเรียงข้อความเหล่านี้ตามลำดับกันอย่างไรจึงจะเหมาะสม ?

1. ก ข ก ง
2. ข ง ก ก
3. ค ก ง ข
4. ง ค ข ก
5. ก ง ค ข

ถ้าบทความที่ยกมานั้นเป็นข้อความเดียวกัน เราก็อาจถามว่าบทความนี้มีลักษณะของการเรียบเรียงเป็นแบบใด

- ก. จากสาเหตุไปสู่ผล
- ข. ให้เหตุผลแบบนิรนัย
- ค. ให้เหตุผลแบบอุปนัย
- ง. จากสิ่งคุ้นเคยไปสู่สิ่งที่ไม่คุ้นเคย ฯลฯ

ทดสอบความแม่นยำหรือความลำเอียงของเรื่องราว เช่นถามว่า

การที่ผู้แต่งกล่าวว่า “.....การตัดสินใจประหารชีวิตเป็นวิธีปฏิบัติแบบป่าเถื่อน” นั้น แสดงว่าผู้แต่งเป็นบุคคลเช่นไร ?... ..หรือจากบทความที่ผู้แต่งเขียนไว้ว่า “หนังสือพิมพ์อเมริกันไม่เคยรายงานความจริงเลย” นั้น ทั้ง ๆ ที่ผู้เขียนไม่เคยเผยแพร่ เราอาจคาดได้ว่าผู้เขียนคือฝ่าย.....

ทดสอบความสามารถในการเก็บใจความสำคัญ เช่นถามว่า

ความสำคัญที่ผู้แต่งพยายามจะเน้นในย่อหน้าแรก คือ — — —
หรือข้อความใดสรุปความเห็นของผู้แต่งได้ดีที่สุด ? — — —

ทดสอบความสามารถในการตีความ

บทความนี้ควรตั้งชื่อว่าอย่างไร ? — — — หรือ เหตุใดผู้แต่งจึงกล่าว
ว่า.....แทนที่จะกล่าว.....

ทดสอบความสามารถในการนำไปใช้ เช่นถามว่า

โดยยึดเหตุผลของผู้แต่งเป็นหลัก เราอาจคาดได้ว่า ความ
พัฒนาการในอนาคตจะเป็นความเจริญทางด้านวิทยาศาสตร์ปานใด? — — —
หรือเหตุผลที่สนับสนุนสหภาพกรรมกรนั้น จะสามารถนำไปสนับสนุน
กฎหมายส่งเสริมการลงทุนในคันไคบ้าง ? — — —

การทดสอบแบบนี้จะใช้ได้ดีมาก ถ้านักเรียนเคยผ่านการอ่าน
แบบ ๆ นี้มาบ้างแล้วในชั้น แต่จะกลายเป็นการทดสอบที่ยุ่งยากอยู่สัก
หน่อย ถ้าความสามารถในการอ่านเก็บความของนักเรียนในชั้นแตกต่างกัน
มาก ฉะนั้น เขาจึงแนะนำให้ใช้บทความอื่น ๆ ที่เด็กยังไม่เคยทำในชั้นมา
เป็นคำถาม แต่ต้องเป็นบทความที่สัมพันธ์กับสิ่งที่เด็กเคยเรียนมาแล้ว
ในชั้นคำถามแบบนี้ใช้กันมากในการสอบแข่งขันชิงทุน เช่น แบบทดสอบ
ชิงทุนการศึกษาแห่งชาติของสหรัฐอเมริกา เป็นต้น

การดำเนินการทดสอบ

เมื่อได้วางแผนการสอบและได้สร้างข้อสอบเสร็จแล้ว ขั้นตอนที่สำคัญ
ต่อไปก็คือการเตรียมดำเนินการทดสอบ ครูส่วนมากมักจะถือว่า
ตอนนี้ไม่สำคัญเหมือนกับตอนอื่น ๆ แต่ความจริงนั้น ถ้าจะให้ข้อทดสอบ

สามารถทำงานตามหน้าที่ที่มันควรจะทำแล้ว ก็จะต้องคำนึงถึงสิ่งแวดล้อมและการเตรียมอื่น ๆ อย่างรอบคอบด้วย ทั้งข้อเตือนใจง่ายๆ เหล่านี้คือ

1. ห้องสอบควรมีแสงสว่างพอเหมาะและมีการถ่ายเทอากาศดี
2. คำชี้แจงที่อธิบายวิธีทำจะต้องแจ่มแจ้งและถูกต้อง
3. ผู้คุมสอบจะต้องทราบว่า แบบ ทดสอบนี้ มีเวลาให้ทำเท่าใด และนักเรียนควรแบ่งเวลาทำอย่างไร
4. แบบทดสอบที่มีลักษณะหรือมีวิธีทำแปลกกว่าธรรมดา จะต้องอธิบายให้เด็กเข้าใจก่อนทุกลักษณะไป
5. จะต้องบอกให้นักเรียนทราบว่า จะตรวจให้คะแนนโดยแบบใด เช่นจะให้คะแนนโดยคิดแต่เพียงข้อที่ตอบถูก หรือจะกิตคะแนนแบบหักข้อผิด เพื่อกันการเกา เป็นต้น
6. ต้องบอกให้นักเรียนทราบด้วยว่า เขาจะต้องตอบโดยเลือกคำตอบที่ดีที่สุดเพียงคำตอบเดียว หรือว่าจะต้องตอบข้อละหลาย ๆ คำตอบ
7. จะต้องขจัดสิ่งรบกวนจากภายนอกหรือการชักจูงระหว่างระหว่าง การทดสอบให้มันน้อยที่สุด
8. ควรจัดที่นั่งให้มีโอกาสเกิดทุจริตได้น้อย ๆ ถ้าจำต้องสอบนักเรียนมาก ๆ ในห้องเรียนเล็ก ๆ
9. ผู้คุมสอบจะต้องเดินไปรอบ ๆ ห้อง คอยจับตาดูว่านักเรียนต่างคนต่างทำโดยตนเอง- แต่จะต้องเดินเบาๆโดยไม่รบกวนและทำลายสมาธิของเด็ก

การตรวจและการวิเคราะห์ผลการสอบ

เมื่อสอบเสร็จแล้ว ควรจะตรวจและแจ้งผลการสอบให้นักเรียนทราบโดยเร็วที่สุด หลักการที่ดีนั้น จะต้องใช้ข้อสอบไม่เพียงแต่เป็นเครื่องมือวัดถ่ายเดียว แต่จะต้องใช้ให้เป็น เครื่องมือการสอนด้วย ถ้าการสอบนั้นใช้กระดาษคำตอบต่างหากอีกแผ่นหนึ่งแล้ว เราก็ควรให้ตัวข้อสอบนั้น แก่เด็ก เพื่อเขาจะได้ไปพิจารณาความผิดพลาดต่าง ๆ ที่เขาทำ (และจะช่วยชี้จุดอ่อนและหรือคำถามสองนัยของข้อสอบด้วย) และค้นหาคำตอบที่ถูกต้องโดยตัวเขาเองได้ด้วย

ครูที่เพิ่งเคยใช้ข้อสอบแบบปรนัยเป็นครั้งแรก อาจจะรู้สึกแปลกใจที่เห็นคะแนนของเด็กกระจายแตกต่างกันมากถ้าแบบทดสอบนั้นมีความยากง่าย ความจำแนกสูง แม้ว่าผู้เชี่ยวชาญในการสร้างแบบทดสอบยังมีความเห็นไม่ลงรอยกัน ว่าคะแนนต้องกระจายปานใดจึงจะสมควร แต่ก็อาจกล่าวเป็นกลางๆไว้ก่อนว่า คะแนนมัธยฐาน ควรจะอยู่ในระหว่าง 50 % ถึง 60 % ของคะแนนเต็ม คราบใดที่ข้อสอบยังมีได้ทดลองใช้หลาย ๆ ครั้ง จนครูสามารถตั้งเกณฑ์ผลการสอบได้อย่างถูกต้องแล้ว ครูก็สมควร ไปกำหนดว่าคะแนนเท่าใดจึงจะได้เกรดใดเป็นอันขาด เมื่อตรวจแล้วก็ควรบันทึกการแจกแจงของคะแนนในกระดาษอื่นอีกแผ่นหนึ่ง ดังในหน้า 51

(จะเห็นว่าคะแนนชุดนี้มีพิสัยกว้างจาก 4 ถึง 19 คะแนน ในข้อสอบที่มีคำถาม 20 ข้อ มัธยฐาน จึงควรอยู่ที่คะแนน 12 แท้ม ซึ่งเป็น 60 % ของคะแนนเต็ม)

คะแนน	จำนวนนักเรียน
20	0
19	1
18	2
17	0
16	2
15	3
14	4
13	4
12	6
11	8
10	4
9	1
8	0
7	6
6	0
5	2
4	1
รวม	44 คน

เมื่อเห็นพิสัยของคะแนนเช่นนั้นแล้ว ครูก็อาจต้องการแปลคะแนนเหล่านี้ให้เป็นเกรดต่อไปอีกชั้นหนึ่ง ขบวนการตัดเกรดนี้ จำจะต้องกระทำโดยดุลยพินิจของครูเอง ว่าเห็นสมควรที่จะให้แต่ละเกรดมีเด็กได้กี่เปอร์เซ็นต์ ก็ครูจะต้องผ่อนปรนตามความยากหรือง่ายของข้อสอบฉบับนั้นด้วย เมื่อกำหนดเปอร์เซ็นต์เด็กในแต่ละเกรดแล้ว ก็ควรจะตรวจดู "รอยขาด" ของคะแนนว่าอยู่ตรงไหนบ้าง ดังในตัวอย่างข้างบนนี้ปรากฏว่า ตรงคะแนน 6 และ 7 ดูที่จะเป็นรอยขาดที่เหมาะสม เพราะเป็นจุดที่แบ่งความสามารถของเด็กได้เด่นชัดมากกว่าที่อื่น ๆ ถ้าเราเห็นว่าสมควรแล้ว (สอดคล้องกับจำนวนเปอร์เซ็นต์ที่เรากำหนดไว้) เราก็จะได้ใช้รอยขาดทั้งสองแห่งนั้นเป็นหลักในการตัดเกรด เช่นสมมุติว่าครูกำหนดจำนวนเปอร์เซ็นต์ของเด็กในแต่ละเกรดไว้ดังนี้ คือ

เกรด A=10 %

เกรด B=20 %

เกรด C=50 %

เกรด D=25 %

เกรด F= 5 %

รอยขาดก็จะตกตรงคะแนน 18 และมากกว่า	(6.8 %)
คะแนน 14 ถึง 17	(20.4 %)
คะแนน 8 ถึง 13	(52.3 %)
คะแนน 7	(13.6 %)
คะแนน 6 และต่ำกว่า	(6.8 %)

หลังจากให้เกรดเสร็จแล้ว ครูก็ควรจะกลับมาที่แบบทดสอบนั้นอีกทีเพื่อตรวจดูว่า ข้อสอบฉบับนั้นบอกถึงความสัมฤทธิ์ผลของเด็กอย่างไรบ้างและเด็กได้บรรลุถึงวัตถุประสงค์ของวิชานั้นปานใด เรื่องนี้จะกลายเป็นของง่ายมาก ถ้าครูได้ทราบ 2 ประการนี้ไว้ก่อน คือ

1. จำนวนเปอร์เซ็นต์ของเด็กที่ตอบคำถามแต่ละข้อผิด หรือที่เรียกว่าความยากของข้อคำถาม หรือ “ดัชนีความยาก” และ

2. ตัวเลขดัชนีที่บอกเราว่า คำถามแต่ละข้อๆ นั้นมีสมรรถภาพในการจำแนกเด็กออกเป็นพวกเด่นที่มีคุณสมบัติสูง (มีความรู้และทักษะสูง) กับพวกค้อยที่มีคุณสมบัติต่ำ เป็นมูลค่าเท่าใด (ค่านี้เรียกว่า “ดัชนีการจำแนก” ซึ่งบอกอำนาจความจำแนกของคำถามแต่ละข้อ) วิธีง่ายๆ สำหรับหาดัชนีการจำแนก ก็โดยแบ่งกระดาษคำตอบนักเรียนทั้งหมดออกเป็น 3 กลุ่ม เท่า ๆ กัน คือแบ่งเป็นพวกคะแนนสูง 1 ส่วน พวกคะแนนปานกลาง 1 ส่วน และคะแนนต่ำ 1 ส่วน แล้วหาคว้ามี่เด็กใน

กลุ่มสูงตอบคำถามแต่ละข้อถูกเป็นสัดส่วนเท่าใด และเด็กในกลุ่มต่ำตอบข้อนั้น ๆ ถูกเป็นสัดส่วนเท่าใด แล้วก็นำมาลบกันได้เป็นดัชนีการจำแนกของข้อนั้น ๆ ถ้ามีนักเรียนสอบน้อย ก็ควรแบ่งคำตอบออกเป็นเพียง 2 กลุ่ม ไม่ควรแบ่งเป็น 3 กลุ่ม

ถ้าเราเอาดัชนีความยากของคำถามมาเรียงกัน ตามลำดับแล้ว ครูผู้สร้างข้อสอบก็จะทราบความยากของคำถามแต่ละข้อ ๆ เหล่านั้นได้ เรื่องนี้มีคุณค่ามาก เพราะจะช่วยให้ครูสามารถจัดข้อสอบให้มีทั้งข้อยากและข้อง่ายก้ำกึ่งกันในโอกาสต่อไปด้วย และยังมีคุณค่าช่วยให้ครูตัดสินใจได้ดีขึ้นว่า ในการสอบครั้งนั้นควรจะให้เกรดนักเรียนถึงหรือหย่อนปานใดอีกด้วย ส่วนดัชนีการจำแนก ก็จะช่วยให้ครูประเมินคำถามแต่ละข้อ ๆ นั้นได้ว่า สามารถทำงานตามที่เรากำลังต้องการหรือไม่ ทั้งนี้ก็โดยสังเกตดูว่า ถ้าเป็นข้อคำถามที่ดี นักเรียนในกลุ่มสูงจะต้องตอบถูกมาก ส่วนนักเรียนกลุ่มต่ำมักจะเลือกตอบตัวลวงหรือตัวผิด ดัชนีทั้งสองนี้มีความสำคัญที่จะให้ครูประเมินประสิทธิภาพของแบบทดสอบ และอาจใช้เป็นข้อมูลสำหรับเปรียบเทียบคำถามแต่ละข้อที่ได้ใช้ในปัจจุบัน กับที่จะถูกนำไปใช้ในอนาคตกด้วย

ถ้าเรายังสามารถทราบคุณภาพของแต่ละตัวเลือกในคำถามทุก ๆ ข้อแล้ว ก็ยังดีและมีประโยชน์สูงขึ้น นี่เป็นงานอีกขั้นหนึ่งเรียกว่า **การวิเคราะห์ข้อทดสอบ** ซึ่งจะช่วยให้ครูทราบว่า นักเรียนทั่วไปยังเข้าใจผิดในเรื่องใดอยู่ และควรจะปรับปรุงการสอนในภายหน้าอย่างไร การวิเคราะห์นี้จะบอกให้เราทราบว่า ตัวลวงใดไม่มีนักเรียนเลือกเลย ซึ่งในการสอบคราวต่อไป เราจะได้เปลี่ยนตัวลวงที่ไม่มีสมรรถภาพเหล่านั้นด้วยตัวลวงอื่น ๆ จนเด็กอ่อนบางคนหลงมาเลือกตัวลวงใหม่นั้นเป็นคำตอบ

ในการวิเคราะห์แบบทดสอบนี้ ยังมีตัวประกอบที่ควรจะต้องกล่าวถึงอีก 2 อย่าง คือความเที่ยงตรงและความเชื่อมั่นของข้อสอบ เรามีวิธีตรวจสอบความเที่ยงตรงของแบบทดสอบได้หลายวิธี เช่นเราอาจตรวจดูว่า แบบทดสอบของเรานั้นมักจะให้คะแนนผลสอบ ใกล้เคียงกับข้อทดสอบฉบับอื่น ๆ ที่เป็นประเภทเดียวกันหรือไม่ หรืออีกวิธีหนึ่งก็โดยพิจารณาจากความมุ่งหมายของการสอบ ว่าข้อสอบฉบับอื่น ๆ นั้นมีคำถามที่วัดทักษะที่เด่นและสำคัญๆ ตรงตามความมุ่งหมายที่กล่าวไว้หรือไม่ การตรวจดูความเที่ยงตรงของข้อสอบ ก็มีอิทธิพลต่อการให้เกรดของครูอยู่เหมือนกัน จากการวิเคราะห์นี้จะสามารถชี้ให้เห็นจุดอ่อนและแผ้วแผ่ต่างๆของการสอบได้เป็นอย่างดี จึงเท่ากับเป็นการปรับปรุงการสอบวัดของเราให้ดียิ่งๆ ขึ้นได้ ในอนาคต

ความเชื่อมั่นของแบบทดสอบ หมายถึงคุณสมบัติที่นักเรียน จะได้คะแนนเท่าเดิม ถ้าได้นำข้อสอบประเภทเดียวกันหรืออีกฉบับที่คล้ายกัน มาให้นักเรียนนั้นสอบซ้ำอีกที (อนุมานว่ายังไม่มีการเรียนรู้เพิ่มเติม ฯลฯ) ความเชื่อมั่นนี้อาจหมายถึงคุณภาพที่แบบทดสอบสามารถแจกแจงได้ออกเป็นประเภทๆ ได้คงที่ตามนั้นทุกครั้งไปได้ ถ้าแบบทดสอบใดไม่สามารถแสดงความคงเส้นคงวาออกมาให้ประจักษ์ แบบทดสอบนั้นก็ไม่นับว่าเป็นเครื่องมือวัดสำหรับวินิจฉัยความสามารถของนักเรียนเลย สำหรับครูทั่วไปนั้น เพียงแต่ใช้พิสัยของคะแนนเป็นเครื่องชี้บอกความเชื่อมั่นก็ควรจะเป็นการเพียงพอแล้ว คือถ้าข้อสอบใดมีพิสัยของคะแนนกว้างก็อนุมานได้ว่ามีความเชื่อมั่น และพึงระลึกไว้อีกประการหนึ่งว่า ถ้าข้อสอบใดไม่มีความเชื่อมั่น ข้อสอบนั้นก็ไม่มี ความเที่ยงตรง

อภิธานศัพท์

(จาก "A Glossary of Measurement Terms," California Test Bureau, Monterey, California.)

แบบทดสอบผลสัมฤทธิ์ (Achievement test) — แบบทดสอบที่ใช้สำหรับวัดระดับความรู้ของนักเรียนเป็นรายวิชาหรือหลายๆ วิชา หรือสำหรับวัดความสามารถทั่วไปของการเรียน

การแก้การเดาคำตอบ — วิธีปรับแก้คะแนนการสอบโดยใช้จำนวนข้อที่ตอบผิดมาประกอบการพิจารณา ทั้งนี้โดยถือหลักว่า ถ้านักเรียนตอบข้อสอบปรนัยโดยการเดาแล้ว จำนวนครั้งที่ตอบถูกและผิดจะเป็นสัดส่วนกับจำนวนตัวเลือกของแต่ละข้อ สูตรทั่วไปที่ใช้ในการปรับแก้ คือ

$$S = R - \frac{W}{n-1}$$

ในเมื่อ S คือคะแนนที่ปรับแก้การเดาแล้ว

R คือจำนวนข้อที่ตอบถูก

W คือจำนวนข้อที่ตอบผิด

n คือจำนวนตัวเลือกของแต่ละข้อ

สัมประสิทธิ์ของสหสัมพันธ์ — มาตรการวัดความสัมพันธ์ระหว่างสองสิ่งสำหรับวัดแนวโน้มของเรื่องราว 2 เรื่อง หรือมากกว่า หรือระหว่างคุณสมบัติใดๆ ที่เกี่ยวข้องกัน สหสัมพันธ์นี้มีค่าจาก -1.00 ซึ่งหมายถึงความสัมพันธ์แบบกลับกันอย่างสมบูรณ์ จนถึง 0.00 ซึ่งไม่สัมพันธ์กันเลย และ +1.00 ซึ่งมีความสัมพันธ์ตามกันอย่างสมบูรณ์

เกณฑ์—มาตรฐาน เกณฑ์ปรกติ คือหลักการวินิจฉัยที่ใช้สำหรับเปรียบเทียบปริมาณหรือคุณภาพต่าง ๆ

ค่าความยาก — จำนวนร้อยละของกลุ่มผู้เข้าสอบที่ตอบคำถามข้อนั้นๆถูกต้อง

อำนาจจำแนก — สมรรถของข้อคำถามที่สามารถแยกบุคคลผู้มีคุณสมบัติไต่สูง (เช่น ทักษะ, ความรู้, หรือทัศนคติ) ออกจากผู้มีคุณสมบัติอื่นๆต่ำ

การแจกแจงความถี่ — การรวบรวมและบันทึกคะแนนตามลำดับจากสูงไปหาต่ำ (หรือจากต่ำไปสูง) เพื่อแสดงจำนวนคนที่ได้แต่ละคะแนนนั้นๆ หรือแต่ละช่วงคะแนนนั้นๆ

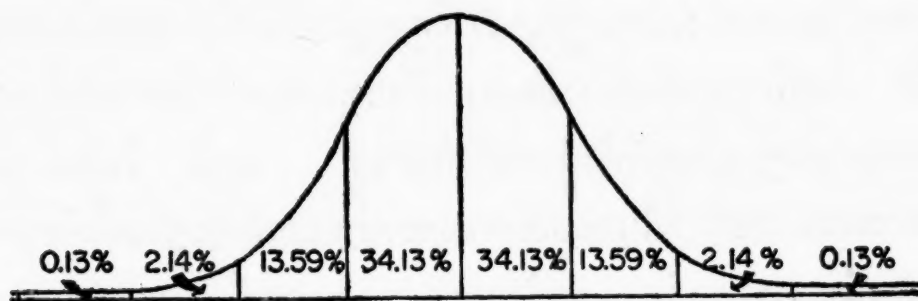
การวิเคราะห์ข้อสอบ — วิธีการอย่างหนึ่งในขบวนการสร้างแบบทดสอบเพื่อจะหาว่า คำถามแต่ละข้อนั้นสามารถจำแนกผู้สอบออกเป็นประเภทๆ ได้ดีปานใด คุณภาพของข้อคำถามขึ้นอยู่กับองค์ประกอบ 3 ประการคือ (1) ความเที่ยงตรงของคำถามที่มีต่อหลักสูตร และความมุ่งหมายของการศึกษา (2) อำนาจจำแนกของข้อคำถามที่มีต่อความเที่ยงตรงและความเชื่อมั่นของแบบทดสอบนั้น (3) ความยากของข้อคำถาม (ดูคำว่าเกณฑ์ประกอบ)

รายเฉลี่ย — ผลรวมของคะแนนทั้งหมดหารด้วยจำนวนครั้ง

มัธยฐาน — คะแนนกึ่งกลางของคะแนนชุดหนึ่งที่เรียงกันตามลำดับมัธยฐานเป็นจุดที่แบ่งคะแนนออกเป็นฝ่ายที่สูงกว่าและต่ำกว่าเป็นจำนวนเท่าๆ กันซึ่งตรงกับเปอร์เซ็นต์ไทล์ที่ 50 พอดี

โค้งปกติ — เส้นโค้งที่ยึดสมมุติฐานว่า การแปรผันจากรายเฉลี่ยเป็นไปตามโดลกบังเอิญ โค้งนี้มีทรงเป็นรูประฆัง ซึ่งเป็นที่ยอมรับกันว่าเป็นความจริงตามนั้นด้วย เพราะมีปรากฏการณ์ยืนยันซ้ำๆ เป็นอันมากว่าการแจกแจงความถี่จากการวัดคุณลักษณะต่างๆ ของมนุษย์ในเชิงจิตวิทยา

และการศึกษานั้นมีทฤษฎีตรงตามนี้ โกงนี้ก็มีคุณค่าในทางคณิตศาสตร์มาก ในโค้งปรกตินั้น คะแนนต่างๆ จะกระจายเป็นสมมาตรจากรายเฉลี่ยเสมอ คือในแต่ละช่วงที่เท่าๆ กันนั้น จะมีจำนวนคะแนนที่สูงกว่าและต่ำกว่ารายเฉลี่ยเป็นจำนวนเท่ากัน และคะแนนส่วนใหญ่จะออกันตรงรายเฉลี่ย แล้วจะค่อยๆ ลดน้อยลงไปทั้งสองข้าง



เปอร์เซ็นต์ไทล์—จุด ๆ หนึ่งของมาตรการวัด ที่แบ่งอันดับของคะแนน ออกเป็น 99 ส่วน โดยแต่ละส่วนมีคะแนนอยู่ 1/100 ราย ตำแหน่ง เปอร์เซ็นต์ไทล์ของบุคคลใด ก็คืออันดับของบุคคลนั้นในกลุ่มมาตรฐาน จำนวน 100 คน ที่เป็นตัวแทนทั้งสิ้นทั้งปวงของประชากรปรกติ ถ้าใคร ได้เปอร์เซ็นต์ไทล์ที่ 70 ก็หมายความว่า เขามีฐานะเท่ากับหรือเหนือกว่า บุคคลอื่น ๆ ในประชากรนั้น 70% หรือจะแปลคะแนนเปอร์เซ็นต์ไทล์ ที่ 70 นั้นว่า ยังมีเด็กทั่วไปอีก 30% เก่งกว่าเขาก็ได้

พิสัย — ผลต่างระหว่างคะแนนสูงสุกกับต่ำสุดจากผลการทดสอบใดๆ ของ กลุ่มนั้น

ความเชื่อมั่น — ระดับความเชื่อถือได้ของคะแนน ว่านักเรียนแต่ละคน จะได้คะแนนเท่าเดิมหรือไม่ ถ้าให้เขาสอบแบบทดสอบนั้นซ้ำอีกครั้ง (อนุมานว่ายังไม่เพิ่มการเรียนรู้ ไม่เกิดการฝึกฝน ฯลฯ)

บ้ายเบนมาตรฐาน — (คูโคงการแจกแจงปรกติ) เป็นหน่วยทางสถิติสำหรับบอกปริมาณความบ้ายเบนจากรายเฉลี่ยของการแจกแจงนั้นๆ ค่านี้หาได้โดยถอความรู้สองของรายเฉลี่ยของความบ้ายเบนทั้งหมดที่เหจากมัชฌิมเลขคณิตยกกำลังสอง ถ้านักเรียนที่สอบเป็นกลุ่มปรกติ เรานำคะแนนจากการสอบนั้นมาจตุกราฟก็จะได้เป็นโคงปรกติ และจะมีจำนวนคะแนนประมาณสองในสาม (68.3%) ที่ตกอยู่ภายในหนึ่งบ้ายเบนมาตรฐานเหนือ และใต้รายเฉลี่ยกับจะมีคะแนนอีกประมาณหนึ่งในสามอยู่ที่อีกหนึ่งบ้ายเบนมาตรฐานที่อยู่ถัดขึ้นไปกับที่ถัดลงมา ฉะนั้น จึงมีคะแนนอยู่ประมาณ 95% ตกอยู่ภายในขอบเขตของสองบ้ายเบนมาตรฐานทั้งที่อยู่เหนือกว่าและต่ำกว่ารายเฉลี่ย และถ้าขยายขอบเขตออกเป็นสามบ้ายเบนมาตรฐานทั้งเหนือกว่าและต่ำกว่ารายเฉลี่ยแล้ว ก็จะครอบคลุมคะแนนถึง 99.7%

ความเที่ยงตรง — ความเที่ยงตรงของแบบทดสอบ ขึ้นอยู่กับที่เราทราบว่าแบบทดสอบนั้นวัดหรือพยากรณ์อะไรได้บ้าง หรือแบบทดสอบนั้นสามารถวัดสิ่งที่เรามุ่งหมายจะวัดได้หรือไม่

I. EVALUATION AND EVALUATIVE INSTRUMENTS

In recent years the purpose and importance of evaluation has taken on a new meaning to the people of Thailand. When the educational system was smaller and the nation was less concerned with national development it was possible to have more expensive (and oftentimes less practical) educational processes. Now, however, with new needs and demands brought by emphasis on new national goals, legislation extending the years of compulsory attendance and the increased number of children to be educated it has become necessary for school teachers and administrators and government leaders to take a long hard look at existing practices to see whether the present educational system is the most economical and efficient organization that can be produced with the available budget. This analysis is done through the process of "evaluation". Good evaluation requires the employment of some measuring devices which can provide evidence which shows what is now being done and which can suggest some ways for improving on the status quo.

The purpose of this brochure is to help educators to better understand the value of good measurement and to offer some suggestions for improving evaluative techniques and instruments. It is based on the assumption that continuous and meaningful measurements must be used for the betterment of teaching, learning, guidance, curriculum planning and administrative efficiency. Such an assumption demands that educators must begin with a very positive attitude about evaluation and that they must be intolerant of bad measuring devices and haphazard construction of examinations. Hopefully, it will dispel the notion held by some teachers that a good test can be put together in a few minutes before it is scheduled to be administered; good testing, like good teaching, demands careful planning and organization.

Evaluation and Objectives

Fundamental to the learning process is a statement of objectives which provide the goals of education and the process

by which these goals are reached. Good objectives, of course, will take into account not only what the teacher intends to do but, particularly, what is expected of the students. Meaningful objectives will be shared with the students so they will know what they are expected to achieve in their work. At the end - or at given points during the instruction period - an evaluation should be made of how well the objectives were achieved. It is reasonable to assume that the goals and the results may not exactly correspond so, realistically, one must expect to measure the "results" and through this analyze how well the original objectives were accomplished. This should not be interpreted to mean that the revision of objectives should be based solely upon student achievement but it does mean that realistic goals must take into account what can be reasonably expected of students to achieve. Gradually, what was initially a long list of objectives that would represent "ideal" accomplishments can, through planned testing, be reduced to a shorter and more realistic list of objectives.

Evaluation and Student Motivation

Another vital role of evaluation is to motivate students. The wise teacher will attempt to encourage self-evaluation as a means of extending student interest and achievement. Classroom tests do much to stimulate students to keep up with and improve their work. No one needs to be reminded that when students are confronted by an impending examination--many are moved to "cram" or review. Knowing that they will be evaluated on their performance usually affects the times when students study and the intensity of their efforts. Properly constructed evaluative instruments can do much to encourage students to develop good study habits and, similarly, badly constructed examinations can confuse, discourage and mislead students.

Examinations and Grades

One of the major responsibilities given to the teacher is the difficult, but necessary, task of assigning grades. Considering the

importance of grades to the students, one must be as certain as he can be that this job is not done whimsically. Grades should reflect both achievement and quality of performance. A high or low grade would be a determining factor in finding a job, being permitted to sit for university entrance examinations, establishing individual interest in certain careers and vocations or as motivation or deterrent for successful performance in school. Examinations help to determine not only the degree of achievement but also the individual's achievement relative to others in his class. For all these reasons, evaluative instruments must be as reliable and valid as they can be because they will provide the basic measurements from which the final grade judgment is made.

Evaluation as a Learning Instrument

Closely allied with motivation is the learning process itself. The educational value of an examination is dependent almost entirely upon its quality. Good examinations have beneficial effects—bad examinations can be harmful. Proper tests, properly used, can help students to correct errors and misunderstanding. They can help identify weaknesses and strengths in the way things are taught and the way things are learned. When failures occur, either by individual students or by the class as a whole, a good test will help to identify the cause for the breakdown. A good test can also provide useful material for teaching; by reviewing the test in class students can see what objectives the teacher had in mind as well as relate the examination experience to the other instruction of the class.

Evaluation and Understanding Students

Obviously, education consists of something more than programs, instruction and learning processes; basically, it is concerned with the enlightenment and fulfillment of human beings. Since this is a basic concern, it is important to know as much as is possible about the individuals who are being educated. Various evaluation instruments are now available which help to measure

the students' aptitudes, attitudes, value systems, potential for academic success and level of experience and achievement. It is only when such things are known that the educational institutions can establish realistic programs, realistic objectives and have an adequate base for the guidance and counseling of students.

Summary

Good evaluation consists of much more than simple achievement tests. Properly planned evaluation which utilizes proper evaluation devices can do much to implement the whole learning process: e.g., in the setting of objectives, serving as a motivation to learning, providing practice in the development of certain associated skills and aiding in the development of new attitudes, providing reinforcement of things learned by showing relationships and organization of knowledge and, hence, promote better understanding and, finally, permitting the student who has performed well to have the satisfaction of a reward through a good grade. Good evaluation not only checks the student but aids the teacher who can profit from the information to improve future teaching.

II. TYPES OF MEASURING INSTRUMENTS

Certainly one of the most difficult tasks that the teacher contends with in his work is the job of evaluating student progress and his own effectiveness. This task becomes less formidable if he is aware of the different types of measuring instruments that are available and the functions that each best serves. There have been many new devices invented in recent years as professional evaluators have attempted to find new methods which will be more accurate and more useful in measuring skills, attitudes, value systems, etc., that were, previously, considered to be beyond the scope of measurement. Our concern here will be with three types which are commonly used and do not require special equipment that may not be available to the teachers of Thailand, i.e., oral and written (free-response and fixed-response).

Oral Tests

In some respects, the oral examination could be the most effective means of really discovering what students know, how they can relate knowledge and experiences and their levels of understanding and sensitivity. Since this type of test demands considerable time and expense other evaluating instruments are used more often, especially when large numbers of students are to be tested. The oral examination has a number of advantages: (1) It brings the examiner and the student into direct contact and makes it easier for the student to be certain of the interpretation of the questions. Not unrarely do teachers find that students misunderstand the meaning of questions on written tests and score badly even though they might have had complete understanding of the material being covered. (2) The oral examination makes it possible for the examiner to probe and pursue certain points to discover levels of understanding — or the student's ability to relate one idea with another. This spontaneous questioning affords greater latitude in examining. (3) This type of examining allows the examiner to observe the student's reaction to questioning. Hesitancy, signs of emotional stress, etc., should help the examiner evaluate the student's competence. (4) In some instances this kind of examination is necessary, e.g., for students with physical handicaps (blindness, loss of muscular control, etc.). (5) Some students react better psychologically when they can perform directly before another person or group of examiners but the good examiner will be wary not to confuse personality with competence. (6) Probably this form of examining corresponds more closely with the ordinary academic and non-academic examining people face in their everyday lives and, hence, seems more natural to the student. Ordinarily, one would expect to be questioned and make replies through the medium of speaking in the conduct of regular living rather than via written responses. The oral examination then is helpful in giving experience to students for this kind of interview relationship. The oral examination is particularly appropriate to students who have been taught using the common teaching methods one finds at all

levels of education in Thailand. Students have relied heavily on the lecture-response learning method and may find the oral test easier.

While there are many advantages—there are also some obvious disadvantages. (1) Probably the biggest disadvantage is the **unreliability** (See Glossary—"Reliability") of this type of test. It is difficult for the examiner to be consistent in judging the performance in these examinations. When the student is examined by a group the usual practice is to take a consensus of the judgments for the final appraisal much in the same way as each judge in a boxing match hands in his decision and the majority opinion wins. Such polled opinion may not be an infallible way of judging—but it probably is the fairest means as far as the student is concerned. (2) A second possible disadvantage is the lack of **validity** (See Glossary) of many oral examinations i.e., they don't really measure what they are supposed to measure. As mentioned above the examiners must be cautious not to be swayed by verbal proficiency (unless this is one of the objectives of the examination) rather than the student's competence and his achievement of the real objectives of the course. Closely tied to this is the tendency for the examiner to be affected by the student's appearance or mannerisms. It is quite possible that a cocky student could offend the examiner and be failed on the basis of this rather than his competence in the field. On the other hand, it is conceivable that a pretty girl might receive some credit from male examiners for her attractive appearance rather than her understanding of the subject.

(3) One often finds in oral examinations that time limitations make it difficult to really cover the field adequately. It is not unusual for students to go into a long explanation of some points and this, coupled with the examiner's attempt to discover the student's depth of understanding, sometimes makes it impossible to cover the breadth of the student's knowledge. Often in an hour examination the time passes so quickly that the examiner doesn't have a great deal of evidence on which to make a judgment of the student's work. (4) Another factor which contributes to

making many oral examinations invalid is that examiners oftentimes fail to plan their sequence of questioning and rely on questions which arise spontaneously. Again, it is important to keep in mind what objectives are to be covered and a carefully thoughtout examination is necessary to keep the discussion from rambling and centered on the most important subjects. (5) A final disadvantage is the great drain on the teacher's time. When separate times are set for individual students to be examined the teacher with a large class finds that it simply isn't worth the effort. When the burden becomes too heavy—the pressure tends to increase the chances that these examinations are invalid and unreliable.

Written Examinations

Written examinations are the most common form of measuring student achievement. They have tended to replace the oral examination in most classrooms because they have the advantage of presenting the same job to every member of the class and larger numbers of students can be examined with the same time limitations. There are several types of written tests: essay, short answer and completion are the ones ordinarily used. As with the oral examination approach, each has some distinct advantages and equally distinct disadvantages and, perhaps, a short review of these would be helpful to the teacher in deciding which approach most nearly satisfies the need of each given situation.

The essay-type test is very popular in Thailand. Teachers find that this form of examining elicits certain responses that one cannot get in other types of written tests. This free-response examination permits the student to reply to a question in a way which reflects his own thinking. Further, it tests a student's ability to organize materials in a systematic way and his skill in expressing himself. Teachers who prefer this type of test over the multiple-choice test argue, logically, that an essay requires the students to SUPPLY the answer rather than simply RECOGNIZE the answer. This means, of course, that the chances for the student to guess are minimized. A properly structured question can draw forth the student's

ability to select important facts or ideas—relate them to one another—and put them into some meaningful context.

As with other forms of measurement there are some noteworthy disadvantages too. Essay examinations are often unreliable instruments because of the kind of responses which are brought out and the difficulty in scoring or grading the answers. As with the oral examination, there is the hazard that verbal fluency is sometimes weighted more than real knowledge of the subject. Unless the question is very specific, there is the problem that every answer must have some degree of correctness. The problem for the test reader to grade consistently is also present in this form of measurement. Many studies on this matter have indicated that grade variations occur depending upon the time of day the paper is read—or whether it's read among the first papers in a set or the last. Forty-seven competent educators were asked to grade the papers of ten students with the following results :

Paper	A	B	C	D	F	
1	0	1	17	13	8	A quick glance by the reader will reveal that these examiners reacted quite differently to the student essays. In paper 8, for example, 2 examiners judged the response to be "excellent" but 3 others judged it to be "unsatisfactory ; 9 thought it was "above average" but 13 others marked it "below average"; 19 examiners considered the paper to be "average".
2	0	1	28	11	7	
3	1	6	33	5	2	
4	1	2	30	8	5	
5	2	10	32	3	0	
6	5	14	24	3	1	
7	0	9	15	3	3	
8	2	9	19	13	3	
9	1	1	9	15	21	
10	1	1	20	14	6	
**						

It is painfully apparent that even with qualified raters there is no assurance that grading practices are reliable. It is because of this kind of discrepancy in grading that school officials

** Data by Dr. Osmond E. Palmer, cited in *Evaluation in the Basic College*, Edited by Paul Dressel, Harper & Bros., New York, N.Y.

and teachers are following the practice of having at least two people grade each paper and take an average of the two scores. It is interesting that other studies having teachers re-read the same papers show that they vary in their judgments almost as markedly as the forty-seven readers mentioned in the chart. Examiners often find that they are sometimes influenced by previous work of a student, personal fatigue, handwriting, grammatical errors and similar distractions.

Another free-response kind of written test is the short-answer type in which the student is expected to answer in a word, a phrase or a sentence as in a completion question. Like the longer essay answer it affords the student the opportunity to produce his own answer rather than recognize the correct answer in a multiple-choice form. This type of test is often used as a quiz or in conjunction with a test that may have one or more longer discussion questions. It is useful to check on the students' knowledge of specific cognitive materials, identification of ideas or things, names, dates, vocabulary, etc. One difficulty which is encountered with this type of test is the evaluation of the wide variety of responses which the questions might elicit and too often there is considerable subjectivity in determining the grade. An obvious advantage is that it can, normally, be scored much more rapidly than the longer essay-type and, because more questions can be asked in a given time period than the essay-type, allows for testing more the breadth than the depth of the student's knowledge.

There are several different kinds of fixed-response tests which will be described in more detail later in this booklet. For the moment, our concern is to assess the advantages and disadvantages of the various general types of tests—their merits and demerits.

Probably the most widely used fixed-response (or "objective") test is the multiple-choice test. The proponents of this type of test claim that its development came because of the weaknesses in other forms of measuring. At issue particularly was the comparative reliability in scoring but the critics also pointed to the

other disadvantages cited in the section dealing with free-response questions. (1) Positively, one of the advantages of the objective test is that it permits a broad coverage of any subject matter. Because the questions can be answered quickly it is possible to ask many more questions. (2) For the teacher who is burdened with a heavy class load this type of test permits reliable scoring by an untrained person provided, of course, the "test corrector" is given a list of correct answers to see if the student's responses are similar to the prescribed correct answers. (3) Objective tests also have the advantage of having all of the students approach a problem or question in the same way because the response cannot vary according to the individual whim or reaction of each student. (4) But probably the greatest advantage to the student is the greater reliability of scoring the examination. This type of examination makes it easier for the examiner to check such things as difficulty and discrimination of items* (see Glossary)—hence the reliability and validity of the tests.

There are some real disadvantages of the multiple-choice test item. (1) If a teacher is primarily concerned with the student's ability to write, or his skill in organizing an extended exposition, or his ability to invent hypotheses or create synthesis, probably this form of testing is less adequate than the essay test. (2) In some subjects in which there is little or no agreement as to what constitutes a right or wrong answer (in aesthetics, for example)—forms of testing other than the objective type usually yield better results. (3) Critics of the fixed-response tests also argue that students sometimes achieve better results than they are entitled to because of the possibility of guessing the right answer among the choices. (4) A fourth disadvantage often cited is that the objective test provides only a definite set of alternatives and, hence, the student is not able to indicate what he really knows about a given subject.

The purpose of this booklet is not to take sides as to which form of testing is best. From the foregoing it should be evident that there are some good and bad points about any kind of testing and that the teacher is obliged to choose the instrument which seems best for the given situation. The individual teacher is in the best position to decide which type best suits the needs and

demands of the particular class or subject. When he knows the strong and weak points of each type he can decide more judiciously.

III. CONSTRUCTING EXAMINATIONS

Regardless of the form of test the teacher chooses to use, some common principles and characteristics will apply to all. As pointed out in the previous section, examinations perform a variety of functions. A primary purpose in testing should be to obtain information about student achievement and his ability to use what he has learned. A second purpose should be to use the results of the test to provide direction for teaching and make future learning more productive. The examiner must be well-grounded in the subject matter so he can ask significant questions and be able to discern acceptable answers. The questions must be clearly stated so the student will know precisely what he is expected to cover in his answer. The examiner should decide which type of testing is most appropriate for the situation at hand. So, before the actual test writing begins, he must take into consideration both the objectives which the course sets out to achieve and the outcomes which will provide evidence of what the students actually achieved. His choice of a type of test will depend partly on the objectives to be measured. A good test should not only show what the student learned - but also his ability to apply this knowledge through certain skills which might have been developed in the learning process. Next, it is a good idea to prepare an outline of the subject matter and the various understandings and abilities that should be tested. It is advisable to prepare a preliminary draft of the test, make any necessary modifications and pass on to a colleague to review before it is administered to the students. It will, of course, be important to consider the time allowance for testing to decide on how many and what kinds of questions will be asked. A final decision must be made concerning the difficulty of items. If the teacher is con-

cerned only with the retention of facts, the questions should demand recall of information. If the material is presented well and the students were attentive - the scores will be high. If, on the other hand, the teacher wants to discriminate between various levels of achievement within the class some of the questions should be very easy, some should be difficult enough for the majority to pass but too difficult for the lowest level and some should be VERY difficult, i. e., only the very best students in class will be able to answer the questions correctly. Later there will be a discussion on the importance of keeping records of the difficulty and discrimination of items. Perhaps now some attempt should be made to show the different techniques which can be used in the preparation of tests along with some suggestions for the use of tests and the subsequent evaluation of the worth of the test.

Oral Examinations

There are a number of ways an oral examination can be used - depending upon the material being tested and the particular purpose of the test. For example, if the situation calls for examining the technical skill of the student a series of questions which call for short answers can be asked. It is, of course, possible for the examiner to ask the student to elaborate on any point to see what comprehension levels the student has. If, on the other hand, the examiner wants to test the student's reaction under stress conditions he can begin by taking a provocative stand and proceed to challenge every statement the student makes. This practice would not be advisable for the teacher who is primarily interested in examining what the student knows and it should be noted that it really requires a highly skilled examiner to make this kind of test valid. Usually in the oral examination the examiner follows a pattern of either moving from the inductive approach (specific to general) or the deductive approach (general to specific). A carefully planned approach would be better than asking questions aimlessly. If the

questioning is done in some systematic and orderly way it will probably help the student to recall more of the relevant knowledge he has on the subject. It is worthwhile to remember that if the primary purpose is to measure the student's knowledge it is necessary to have the proper environment. This means that the student shouldn't feel threatened but should be put at ease. Often it helps if a few minutes are allowed at the beginning of the examination for casual conversation.

In organizing an oral examination careful thought should be given to what kind of information is to be sought in the exam, i.e., what questions should be asked and how the responses are to be evaluated. The examiner probably should have an outline of questions and some notion of the relative weight that would be assigned to each part of the examination. Some means must be devised for making an evaluation of each part as the examination is in progress. In cases where the examination is administered by a single examiner, it might be advisable to have a checklist indicating the performance of the student on given sections of the test. If a permanent record is deemed advisable or necessary (or the examiner would like a chance to review what was said) the conversation can be recorded on a tape-recorder. With this kind of examining it is sometimes worthwhile to compare the results with the student's performance on written tests, his classwork, etc. If there is a wide discrepancy in the results of the oral examination and the student's other work perhaps it would be well to weigh the effectiveness of the test or call the student back for further questioning.

Above all, the examiner must keep in mind his purpose in giving the examination. If the examiner takes on a role other than finding out what the student knows he may distract the student from performing at his best. This means that the examiner should avoid the temptation to supplement the information supplied by the student unless this is done intentionally to aid the student

and enable him to recall things that may not seem relevant at the moment. Another potential danger which the examiner must keep in mind is probing too deeply on any given subject so that the test reveals only the student's depth in a given area and does not get out the breadth of his knowledge.

Written Examinations: THE ESSAY.

In the previous section we discussed some of the shortcomings of the essay examinations emphasizing the difficulty of grading or scoring the tests. This should not be construed to mean that they should not be used. On the contrary they serve as very effective instruments especially in courses which include among the objectives the ability to organize, to write creatively, the ability to relate complex ideas, etc. Some of the weaknesses stem from the way the questions are stated, i. e., often they are too general and comprehensive on one hand so that it is difficult for the student to know precisely what the teacher is asking -- or too specific and detailed so that it becomes almost like a fill-in type of test. Ideally, a question should be precise enough so that the student knows what the teacher is seeking and should contain enough suggestions to provide the student with some framework for organizing his answer.

In preparing questions for an essay test consideration should be given to testing over the range of curriculum content covered in the course rather than the details of one segment of the term's work. This helps to improve the reliability of the test because it tests the overall achievement of the students rather than specialization in one given area. Such an approach demands that all students should be asked to answer all questions on the test. Often it is a good idea to begin with a chart which puts the subject matter divisions on one axis and the desired skills or abilities on the other axis.

Subject Matter	Ayudhya		Reigns of	Reigns of	Reigns of	Revolution of	Total
Objectives	Period	Rama I-Rama IV	Rama V-Rama VII			1932 – Present	
Knowledge	1	2	1			2	6
Understanding - Comprehension	3	1	2			1	7
Application	1	1	2			3	7
Judgment	2	1	1			1	5
TOTAL	7	5	6			7	25

By using this type of chart it is possible to see where the greatest emphasis was placed in the instruction and what should be emphasized in the examination. By analyzing the questions the teacher proposes to use in the examination it is possible to see how well the scope of work during the term is being covered.

Before this chart is prepared, it is necessary for the teacher to set down the primary objectives of the course. If, for example, a primary concern is to gain **knowledge** it would be well to state specifically what kind of knowledge: specific knowledge, abstractions, universal knowledge, etc. If an objective is **comprehension** — does this mean understanding for translation or interpreting? **Application, critical thinking** and **evaluating** all might be part of the objectives of the course. Whatever they are, it is important for the teacher to consider these in the questions to see how well the objectives of the course have been achieved.

In writing the questions themselves the teacher must be careful about the wording so the student will know exactly what his task should be. If the question reads; "Tell all you know about political absolutism" — or "Discuss the development of the scientific method" a student can write almost any kind of answer and be correct. Obviously, these questions are too general. If, on the other hand, the questions were to read, "Describe the impact of French colonialism on the foreign policies of King Rama V"—some fairly specific points would be raised which can be evaluated with some degree of reliability. Certain key words can be used to help to make the student's task more specific: Define, Select, Summarize, Compare, — Arrange, etc. They also help to relate the material to the objectives of the course and planning the examination should take this into consideration.

Let's look at some other examples of essay-type questions and what might be done to improve them to measure the achievement of the students and the class objectives.

Ambiguous: "Describe the origins of democracy in Thailand."
In this type of question the student is given no real direction.

What is meant by "Describe"? Does this call for a listing of people who contributed to the democratic processes? Should the response include the student's evaluation of the most important factors in the development of democracy? What is meant by "origins"? Should one begin with the ideas of Phaya Bhahol— or even go back several years to discuss earlier pro-democratic beliefs of benevolent monarchs prior to the Revolution of 1932? What is meant by "democracy"? Political democracy? Social equality? Economic democracy? etc. It would be much easier to elicit a more common response from the students if the question were to read: "Summarize the contributions of Phaya Bhahol to the democratic ideals held in present day Thailand." — Or "In reference to political development, how do the views of Phaya Bhahol differ from those held by earlier political leaders."

Vagueness: "Discuss the effects of the atomic bomb." This could be improved by making the question more specific: "Describe the immediate and long-range effects of the atomic bomb at Hiroshima, taking into account the effect on buildings, people and changing attitudes toward atomic weapons."

In classifying examination questions in accordance with objectives perhaps the following examples will help:

Recall of knowledge:

Simple recall:

"Who was the Thai leader at the Bang Rachan battle?"

Selective recall:

"What four important main international problems between 1890 and 1913 contributed directly to the outbreak of World War I?"

Understanding — Comprehension

By comparison: "What are the physical differences between a mammal and an amphibious animal?"

"What distinguishes a democratic from a republican form of government?"

Explanation of the meaning of a phrase or statement:

"What is meant by the phrase, 'Know thyself'?"

"What is meant by the admonition, 'Lay not up for yourself treasures on earth where moth and rust doth corrupt, and thieves break through and steal. But lay up for yourself treasures in Heaven where moth and rust doth not corrupt and thieves do not break through and steal'?"

Summary or synthesis of something studied in detail:

"What does Darwin mean by 'natural selection'?"

Application: If one of the primary objectives is to have students be able to put their knowledge to use in some new manner other than recalling -- or comprehension -- there are several approaches which can be used. It should, perhaps, be noted that there is a narrow distinction between this type and the "understanding" type mentioned before. Comprehension implies the ability to recall and explain some fact or principle -- application involves having the knowledge, understanding and being able to put this knowledge to use when no solution is specified. Several different methods may be used to test the student's ability to apply his knowledge:

1 - Ability to arrive at generalizations and conclusions to actual living problems:

"In view of the proposed expansion program of the comprehensive secondary schools of Thailand, what changes can be expected in the processes of testing and evaluation?"

2 - Analysis and comparisons:

"Why do present day school teachers prefer written tests to oral tests?"

3 - Ability to relate ideas and/or concepts to another situation:

"How does Buddhism influence the teaching and curriculum in Thai schools?"

4 - Illustration of principles:

"How can one prove scientifically that fresh milk is turned sour by bacteria?"

Judgment: Another objective often found in courses is the ability to evaluate or make judgments. Students should be taught how to recognize value, worth, accuracy, biases, etc. The ability to make judgments, as opposed to opinions, which here we consider to be decisions based upon less substantial evidence, is necessary if a student is to have a deep understanding of his subject. Judgment questions, again, can be approached in several ways:

1 - Decisions regarding right and wrong. -- positive or negative reactions:

"What validity is there to the argument 'Man is the measure'?"

"Should capital punishment be abolished? Why?"

2 - Recognition of accuracy and biases:

"The author of the textbook doesn't like politics. How can you tell even though he doesn't say so directly?"

"German statesmen blame the British for starting World War I, Justify their claims."

Grading the essay examination presents a difficult chore for the teacher largely because of the subjectivity of the student's answers to the questions and the subjectivity of the teacher in reading the student's responses. To increase the reliability of scoring there are a few points to remember. If there are more than one question in the examination -- it is always best to grade and score one question in all the papers then go back and grade, the second, the third, and so on. The reasons for this are obvious --: When

the first answer is graded at a particular level -- there is a tendency for this to influence the grades on the remaining parts of the test. Second, it is too much to expect the grader to make comparisons of quality on particular questions if they are not all read at the same time. Sometimes it helps to read through the papers and separate them into three or four groups according to their general worth: e. g., Superior - Good - Satisfactory - Poor or Unsatisfactory. If a numerical score or letter grade must be assigned then the papers in each group can be reread and scored on a comparative basis. To make the grading as objective as possible, it is a good idea never to have the reader look at the name of the student; this sometimes influences the grade more than the content of the examination.

In grading examinations of this type, one way to improve the consistency of the examiner is by setting up an outline of things which the teacher would expect to have elicited in the responses of the students and predetermine the relative weights which should be assigned to the responses the teacher considers most important. This avoids the possibility that a grade would be assigned by a general impression. A simple way is to write a model examination to outline the points which ought to be covered. Another way is to set up a chart showing points to be covered and their value:

Question: Why did Captain Nora Singh willingly accept death?

To uphold the sanctity of the law	10 points
To demonstrate his respect for and loyalty to the King	10 points
To show his unwillingness to be an exception to the law	10 points
To show his fearlessness of death	9 points
Other reasons	8 points
Organization	8 points
Clarity of expression	8 points
Incorrect statements	Deduct 10 points each

OBJECTIVE TESTS

The objective type test is a more recent development among the measuring instruments. Initially, they were devised to improve on the shortcomings of the oral and essay examinations, i.e., the problem of unreliability in scoring, the opportunities for "bluffing", etc. Whether or not these tests have actually been an improvement as measuring instruments has never been universally agreed upon. The supporters of the objective tests insist that they can prove that their instruments are more valid and reliable. The proponents of the essay tests charge that objective testing encourages memorization of factual details and fails to deal with the most important objectives of education. In all probability, poor test construction has undoubtedly, done more than anything else to arouse criticism of the objective tests along with a lack of proper planning, poorly written test items and badly chosen subjects for test items. If this assumption is true, some simple guidelines and rules should help the teacher to improve this form of test.

The reader will notice that much more time and space will be devoted to the objective test. This is done for two reasons: First, because most teachers have not had the opportunity to investigate the possibilities of this form of testing and, second, there are many more varieties of approaches in objective testing.

Of all the possibilities of test item types within the category of fixed-response questions, probably the most common form which can be applied generally to other types we'll discuss, is the multiple-choice item. The simplest form of this item has two basic parts: the STEM, which presents the problem or begins the statement to be completed—and the RESPONSES which include the correct answer and three or four FOILS or DISTRACTERS. The teacher who is just beginning to experiment with the multiple-choice item would probably be well-advised to start with the items which ask a question—rather than the completion of a statement type. This is simply because there are fewer opportunities to make mistakes with the question-type item. After the question has been posed, the alternatives are offered. The number of responses

depends upon the test writer but, usually, items with four or five responses will be better than two or three response items simply because the former makes it less possible to "guess" the right answer. Some questions do not lend themselves to more than three reasonable responses so it would be preferable to use some other type of item. Ordinarily, it is a good idea to have a consistent number of responses for all the multiple-choice items appearing in the test.

The guidelines or suggestions that follow have special reference to the simple multiple-choice examination but would also apply to most of the other forms which we'll discuss later on.*

(1) The format of the test should be simple and consistent. There is no special advantage in making constant changes in the type of item presented to the student which would require the reading of new directions. A great many variations of items can be reduced to the multiple-choice form. This does not mean that approaches and item styles should NEVER vary. It does mean that there should be a minimum amount of variation within the test and the style should not change unless the particular objectives demand a change. A primary concern of the teacher should be to **AVOID CONFUSING THE STUDENT**.

(2) The nature of the problem should be presented in the stem of an item so that the student will know what he is to look for as he reads the completion. This is true whether the item consists of an incomplete statement followed by several alternative completions or appears in the question form. Example:

Bad Stem: "Which of the following is true?"

Improvement: "Which of the following best describes Gandhi's position on violence?"

(3) The responses should make sense grammatically when each is related to the stem. This is especially necessary when the stem is stated in incomplete sentence form. Generally speaking, neither the incomplete sentence form nor the question form of

* These points are adapted from a list enumerated in "Critical Thinking in Social Science," by Paul Dressel and Lewis Mayhew.

stem has any significant advantage over the other. The best practice is to use whatever form is appropriate to the concept and will avoid confusion and unnecessary wordage.

(4) Items should not confuse the student by use of excess verbiage or by intricate sentence structure. Example :

Bad item : "During the early 19th century factory owners in England used many devices to further production at low costs to themselves and at the expense of the workers in the factories. Which of the following would provide evidence that this was so ?"

Improvement : "Which of the following best describes the abuses heaped on English factory workers in the early 19th century ?"

(5) The language used in an item should be appropriate to the educational level at which it is expected to be used. For example, the general vocabulary should not be more difficult than the students may be reasonably expected to possess ; otherwise, item difficulty may be increased by irrelevant factors.

(6) Grammatical clues should be eliminated. These often occur when the stem requires a plural answer and only one plural response is given or when clues are given by using "a" or "an" at the end of the stem. Example :

Bad item : When students want general information about a subject they should consult an
 A. encyclopedia
 B. monograph
 C. dictionary
 D. lexicon
 E. newspaper

Improvement : When students want general information about a subject they should consult
 A. an encyclopedia
 B. a monograph
 C. a dictionary
 D. a lexicon
 E. an atlas.

(7) The positions of the correct responses should be varied so that the student cannot profit from the fact that they are often included in the first two responses—or the last response—or that they follow a regular pattern. Before the test is put in its final form it is a good idea for the teacher to record the answers on a sheet of paper to see whether there is any discernable pattern (e.g., B-B-B-D-D-D-A-A-A, etc.) and that there is a fairly good balance of correct answers among all of the responses (e.g., approximately an equal number of A responses—B responses, etc.)

(8) The responses should be homogeneous and plausible—but not to the extent that the incorrect ones contain sufficient truth to allow students legitimately to select one of them as the correct answer.

(9) The concept or objective being tested ought to lend itself to the production of an adequate number of “distracters” or “foils” (wrong answers). If only one wrong answer is thought to be effective, it is a mistake to offer several elaborations of the same theme. If this is done, the student can easily pick out the one answer which is different from the others. Example:

Bad item: Which of the following would you expect to find in a WAT?

- A. The Bot
- B. The Dome
- C. The Mosque
- D. The Spirits' House
- E. The Cross

Improvement: All but one of the following might be found in a WAT. Select the EXCEPTION.

- A. The Bot
- B. The Chedi
- C. The Ho Trai
- D. The Mosque
- E. The Sala

(10) Right and wrong responses should appear to be similar in the degree of precision with which they are stated.

When the right answer is elaborated by the presence of qualifying phrases, the same treatment should be accorded the distracters. If the distracters in such a question consist of brief and careless generalities, the student is quick to recognize that the most carefully written statement is probably the correct one. A useful rule of thumb is to make the distracters of approximately the same length as the correct answer.

(11) Each item should test some significant concept. Items intended to be difficult should not be so because they test for some obscure and probably unimportant information, but because the student is obliged to assess a number of critical elements and then to apply these to a number of carefully selected responses.

(12) Avoid clues which give away the answer in the responses by using words such as "always", "only", "every", etc. It should be noted, however, that OCCASIONALLY these words can be useful in catching the "test-wise" student who would avoid this response simply because this is an obvious attempt to get them to pick the wrong answer.

(13) Items containing double negatives should not be used since they are a source of confusion, i.e., if a word such as "not" or "false" appears in the stem, avoid using another negative word either in the stem or in any of the distracters. It is sometimes well to group together items in which simple negatives appear in the stems. Or the negative may be avoided simply by restating the stem to read: "All but one of the following are accurate statements concerning Renaissance painting. Select the EXCEPTION". or "All of the following are accurate statements concerning Renaissance painting EXCEPT..."

(14) "None of these" or "None of the above" are legitimate responses and probably have the merit of requiring the student to analyze the other responses with particular care. In the test as a whole, the "none" response should sometimes be correct --and sometimes incorrect.

(15) "All of these" or "All of the above" may also be used as a response--but has the weakness of providing a clue when

the student discovers that two of the other responses apply. "All of these" used in conjunction with "None of these" as distracters in the same item has considerable merit. Two main pitfalls should be watched using the response "All of these": If it is used with "None of these" it must precede the "None..." response; also, in the test directions, it must be emphasized that the BEST or most COMPREHENSIVE response is to be selected otherwise students will be able to argue that one of the other responses is correct.

(16) Items should be reviewed by one or more colleagues or student assistants in order to ascertain their clarity, suitability and accuracy.

(17) Items should be appropriately distributed among the various categories of skills which a test or exercise purports to measure.

(18) Ordinarily, the items in a test should range in difficulty from fairly easy to fairly difficult, with the preponderance of items falling near the point of average difficulty.

(19) The items in the test should be carefully scrutinized to make certain that the information given in one item does not intentionally provide a clue for answering another.

(20) The length of the test should be calculated to allow most students to finish in the time allowed.

(21) The test or exercise should be reviewed for punctuation, grammatical and typographical errors. Even more so than in textual materials a typographical error can have crucial significance with respect to the meaning intended and the meaning inferred.

(22) After a test has been given to a class, items should be revised on the basis of a study of the student responses. Class discussion is also helpful in improving the test material for future use.

As with the oral or essay type of test, there are some necessary preliminary steps before one begins the actual writing of test questions. One should not begin simply to WRITE test items and continue until he feels that there are enough to occupy the students' time for the examination period. One should begin

with a chart similar to the one shown on page 73 to make certain that there is a balance in covering the course materials and the course objectives. It should be noted that oftentimes items can be classified under any one of several objectives so this chart can do no more than provide a general outline.

In the make-up of the test some basic practices should be observed. If the test is to consist of different types of test items (true-false, multiple-choice, etc.) it is usually a good idea to group similar types together to lessen the confusion for the student. It is also a good idea to follow a consistent pattern concerning the number of responses per item ; i.e., generally students, in a given test, should be asked to pick the "one best" answer or multiple responses in each item. There is often difficulty in the "multiple correct response" items simply because any wrong responses must be obviously wrong and all correct responses must be unequivocally right. Psychologically, it is best to begin with easier items and make them progressively more difficult as the students progress through the test.

Because of the importance of careful wording in the objective tests it is always advisable to put the test together several days prior to the giving of the test. This enables the author to look at the test items (or better yet, have a colleague review the test) to eliminate ambiguities, unfair questions, accuracy of responses, etc.

True-False Tests

This type of test has been used frequently by teachers simply because it is relatively easy to construct and easy to grade. This type of test has several short-comings primarily because the questions call for one of two answers which is not enough choice to prevent higher scores because of guessing. A second difficulty the test writer encounters is to find statements which are not perfectly obvious which are absolutely true or false—right or wrong. The principle of discerning what is right or wrong, accurate or

inaccurate, is often one of the major objectives of courses; so what needs to be done is to find some means of putting rigor into the true-false test. This can be done if one follows the procedure of using a key list approach:

Example :

Items 1 through 5 with the teachings of the Lord Buddha. Each item consists of a STATEMENT and a REASON. For each item select from the key the relationship that exists between the STATEMENT and the REASON.

KEY :

- A-Both the STATEMENT and REASON are **true**, but they have no cause-effect relationship.
- B-Both the STATEMENT and REASON are **true**, and they are related as cause and effect.
- C-The STATEMENT is **true**, but the REASON is **false** (and does not support the statement).
- D-The STATEMENT is **false**; the REASON is **true** (but not relevant).
- E-Both the STATEMENT and the REASON are false.

Suggestion : First determine whether the STATEMENT is true or false—next determine whether the REASON is true or false—then classify the item according to the key.

1. STATEMENT : Absolute truth does exist.

REASON : This can be proved through sensory impressions.

5. STATEMENT : A true philosopher ought to fear death

REASON : Because he understands that there can be no life-after-death.

This, obviously, would be too complicated for children in the elementary grades but should work well for the pupils in the secondary and post-secondary schools. There are other ways of

using the old true-false item as a measurement of knowledge and eliminate guessing. The teacher can use responses which offer alternatives which would correct incorrect words:

Example: Shale is a sedimentary rock highest in calcite.

The statement is

- A. accurate as it stands.
- B. inaccurate, but would be correct if "slate" were substituted for "shale".
- C. inaccurate, but would be correct if "igneous" were substituted for "sedimentary".
- D. inaccurate, but would be correct if "clay" were substituted for "calcite".
- E. inaccurate, and none of the substitutions makes it correct.

Again, the advantage in this form is that it requires students to pick from five responses and lessens the chance of guessing the correct answer.

A modification of this would be to have the students identify true and false statements and ask that they indicate WHY the statement is correct or incorrect. Sometimes it lessens the confusion on the part of students if key words in the statement are underlined to keep their attention on the significant parts of the statement. If the teacher insists upon using the traditional true-false item there are some special rules which might help to make this a more reliable measuring instrument. Be particularly careful about wording because it is easy to make true-false items "tricky."

If more than a single thought is presented in the item, make certain that both (all) thoughts are positively true or false. To avoid confusion, usually it is a good idea to have a single idea to make certain that the student is being tested for his knowledge rather than his ability to read skillfully.

Examples:

Bad item: "The new constitution in Thailand calls for a two-house legislature whereas there was no provision for a legislative body in the previous constitution".

Improvement: "The new constitution in Thailand provides for a two-house legislature".

(2) Certain words and phrases tend to "give away" the answer. For example, if the statement includes words such as "never"—"always"—"no"—"all", etc., they usually provide the student with a clue that the statement is false. On the other hand, statements which include words such as "usually"—"customarily", "ordinarily" etc., tend to signal a true statement. If such words are used—they ought to be used to disadvantage the student who relies on clues for determining his responses.

Examples :

Bad item : "The King of Thailand is the only official required by law to be a Buddhist."

Improvement : "The Constitution of Thailand stipulates that all monarchs must be Buddhists."

(3) The test writer must be especially conscious of the use of ambiguous words or phrases to avoid misleading the students. When students are confronted with the task of trying to figure out the meaning of certain words—they will lose their ability to demonstrate their knowledge. This particularly applies to adjectives and adverbs which describe degrees of variation : e.g., "often"—"rarely"—"occasionally"—etc. These things will evoke different reactions from different students. How often is rarely ?

Examples

Bad item : "Rarely does one see a humming bird."

Improvement : "Humming birds are less common at the North Pole than they are at the Equator."

Matching or classification items :

A second common form of the objective test concerns **matching** and **classification**. This type of test is best adapted to the association of names, dates, terms, concepts, etc. Actually, the principle in the matching test can be applied to items based upon graphs, maps, pictures or diagrams but, ordinarily, the test

takes the form of matching things in one column with related things in a second column. Probably the most common error in the use of these tests is to use long lists of heterogeneous terms in an effort to eliminate, or at least lessen, the guessing factor. For example :

Column I	Column II
a. Thanom Kittikachorn	1. Newly appointed U.S. Ambassador to the United Nations
b. Dwight Eisenhower	2. French leader who threatened to desert the SEATO organization
c. Ethiopia	3. Revolutionary who brought a communist regime to his country
d. Fidel Castro	
e. Vatican City	
f. Charles de Gaulle	
g. Arthur Goldberg	
h. Prabhas Charusthira	
i. Kasetsart	
j. Moscow	
k. SEATO	
l. Carlos Romulos	
m. Gibraltar	
n. New York	
o. Suez Canal	

One must exercise great care in this type of item to make certain that the student is being tested on his knowledge of the subject rather than his ability to answer through the process of elimination. Note, for example, item 1 (Column II): The student is tipped off by "newly appointed" which indicates that this would have been in the newspapers recently; he is given an additional clue by "U.S."—which makes only two possible answers. It doesn't take much knowledge to pick the correct answer. In item 2 the same kind of clues are given, i.e., "French leader" and "de Gaulle" is the only Frenchman included in the list in Column I. Item 3 also permits the student to go right down the list and by the process of elimination pick the correct answer.

The matching type question can be very effective if a few simple rules are remembered :

(1) It is a good policy to limit the number of items in each of the columns so that there are five **HOMOGENEOUS** items in one column and three, four or six related items in the other. The number of answer choices should not be the same as the number of problems because this will prevent students from getting the right answer through the process of elimination.

(2) Students should be instructed as to whether an answer can be used more than once. If they can be used more than once it is all right to have the same number of answers and problems.

(3) An effort should be made to keep the problems stated in the simplest possible way to keep from confusing the student. An improved version of the matching test might look something like this :

Column I	Column II
A. Sukhothai Period	1. The discovery of Buddha's footprint at Sara Buri
B. Ayudhya Period	2. The abolition of slavery
C. Dhonburi Period	3. The establishmant of Chulalongkorn University
D. Ratana Koesinth Period (before the Revolution)	
E. Ratana Koesinth Period (after the Revolution)	

Another type of objective test similar to the "matching test" is the "key list" or "classification" test. In the key it is possible to use time periods, generalization of concepts or ideas, terms, events, persons, etc. Then following are the ideas, concepts or problems which can be related to the things listed in the key.

Example :

Match the operational services below with the appropriate government agency or office listed in the key.

- KEY:**
- A. The management of the Ampur
 - B. The management of the Changwad
 - C. Either the management of the Ampur or the Changwad
 - D. The cooperative management of the Ampur and the Changwad
 - E. Not under the management of either the Ampur or the Changwad

- 1. Marriage
- 2. Road construction
- 3. Foreign diplomacy

Still another way of testing the student's ability to relate events, men, etc., to historical periods is through the use of a time line.

Example :

In the line below the numbers represent intervals of time before, after and between the historical periods identified in the key. By reference to the key identify the appropriate time period for the events in items 1 through 5.

1.	:	2.	:	3.	:	4.	:	5.
Revolutionary		Civil		Spanish-		World War I		
War		War		American				
				War				

- 1. The French Revolution
- 2. Rama V became King
- 3. Shakespeare wrote **Othello**
- 4. Mussolini became a fascist dictator
- 5. Marx wrote the **Communist Manifesto**

The same principle can be used with a key list if you wish to check the student's awareness of time relationships, chronology or time sequence.

Example :

In items 1 through 4 mark answer space

A—if the event took place before B.E. 1500.

B—if the event took place between B.E. 1500 and B.E. 1800.

C—if the event took place between B.E. 1800 and B.E. 2200.

D—if the event took place between B.E. 2200 and B.E. 2500.

E—if the event took place after B.E. 2500.

1. Abdication of Prachathipok (1935)

2. Napoleon Bonaparte became emperor of France

3. Woodrow Wilson elected President for the first time etc.

For checking the student's knowledge of chronology—it is possible to set up a slightly different key list. Example :

A. Phra Mongkut Kleow

B. Phra Nung Kleow

C. Phra Bhumiphol

D. Phra Lerd Lha Napalai

E. Phra Chulachom Kleow

Then set up the responses :

1. ABCED 2. DBEAC 3. CDEBA 4. BCDEA 5. EDACB

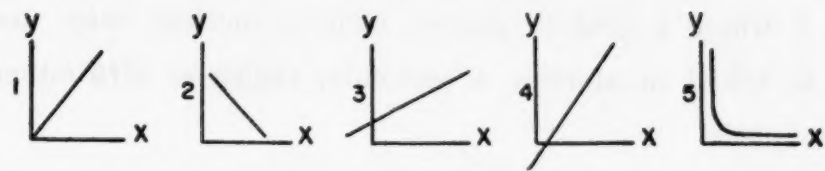
This is not the best form because it does permit the student to eliminate some possible answers if, for example, he is familiar with the first and last person (or other things) in the key list.

Use of graphs and other illustrative material :

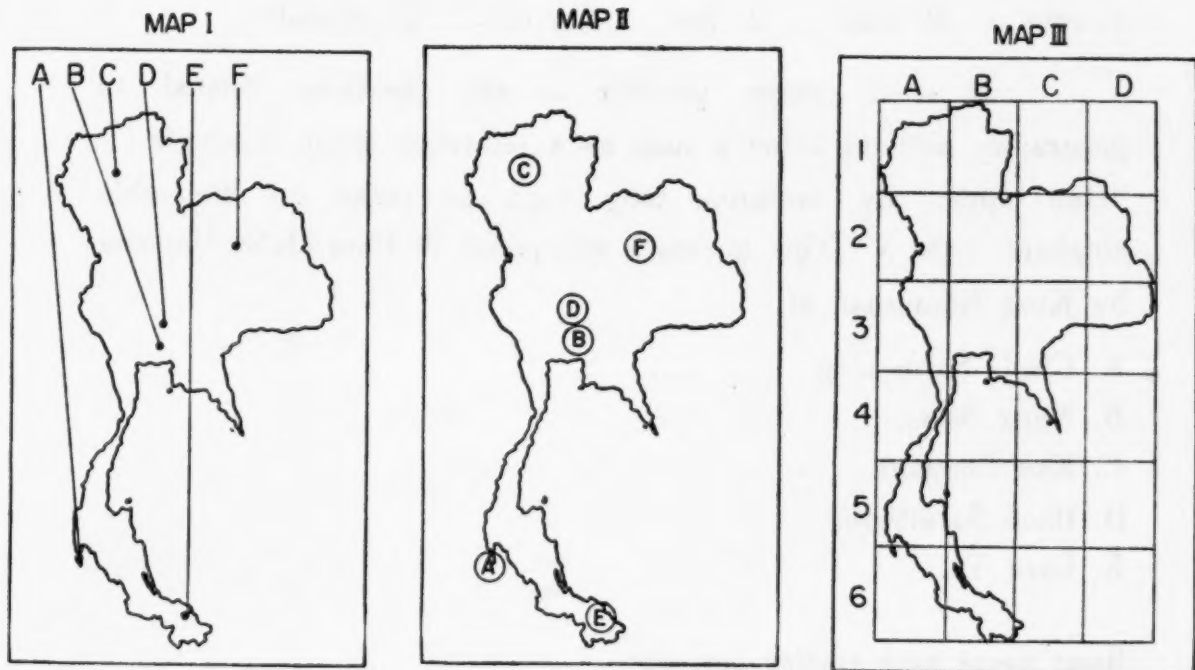
In items 1 through 5

identify the following graphs as to the type of relationship they represent.

KEY : 1-Direct relationship only
 2-Direct proportion
 3-Inverse relationship only
 4-Inverse proportion
 5-None of the above



If knowledge of geography is to be tested, there are several ways that questions can be related to a map. Example I show how particular locations can be identified by use of arrows or lines.



The student can be asked to locate a specific place: Which of the makers designates-or the student might be asked to locate a place which is noteworthy for some important event or landmark:

The fighting which accompanied the attempted coup of Prince Boworadet took place at 1.A 2.B 3.C 4.D 5.F

The same types of questions could be asked if Map 2 is used.

Sometimes the teacher may want to use the style suggested in Map 3 where a grid is placed over an outline map and the student is asked to identify a particular **region** or **area** rather than a specific place.

In which of the following areas would you be most likely to find Hill Tribes ?

1. 1A 2. 2B 3. 3C 4. 4D 5. None of the above

-or Area 3B is most noted for its production of

1. rice 2. teak 3. tin 4. fruit 5. peanuts

It is, of course, possible to ask questions related to geography without using a map as a reference point. Example :
 "How quiet my brethren, why don't we make an honorable elephant fight ?" This question was posed to Phra Maha Upracha by King Naraesuan at

- A. Chedi Sarm Ong.
- B. Nong Sarai.
- C. Kao Chonkai.
- D. Bhoe Sarmthon.
- E. Lard Ya.

Items based upon reading passages :

Another common form used in objective tests concerns measuring the student's skill in reading plus his ability to analyze, make judgments, apply his knowledge, etc. Depending on the purpose for which it is being used, this type of test requires the

selection of a passage which is designed to really test the purpose. For example, if the purpose is to test speed-reading and comprehension a longer selection might be chosen. If the purpose is to test ability to analyze, synthesize, etc., the selection would be shorter. The purpose will also determine how much time should be allowed to enable the student to demonstrate his skills or knowledge of both.

Many things can be done with this type of testing and perhaps a few examples will show their versatility.

Ability to organize or recognize sequence :

One can use three or four paragraphs which are lettered for identification. Then the question could be asked :

Which of the following is the logical sequence in which the paragraphs should appear ?

1. ABCD 2. BDAC 3. CADB 4. DCBA 5. ADCB

— or if the passage is a single essay the question could be asked :
This essay is organized by

- A. cause to effect.
- B. deductive logic.
- C. inductive logic.
- D. familiar to unfamiliar, etc.

Ability to recognize accuracy of evidence or biases :

The author's statement that "...to sentence one to death is a barbarian practice" indicates that he.

— or

According to his passage, American newspapers NEVER report the truth. Although the author's name is not divulged it can be assumed that he is a.

Ability to restate points in the passage :

The main point the author is attempting to make in paragraph 1 is that

- or

Which of the following best summarizes the author's point of view?

Ability to interpret :

What would be an appropriate title for this passage?

- or

Why does the author say.....rather than.....

Ability to apply :

Using the argument made by this author, what can we expect future developments to be in scientific development?

-or

How could the arguments used to support labor unions be applied to support laws to aid management?

This form of testing can be effectively used if there was some reading which was used in class but was fairly involved and complex, especially if there was some difference in the rate or amount of comprehension which the students derived when it was covered in class. Often it is advisable to use some reading passage which is related to something the students had studied but they probably had not encountered before. This form is widely used in certain scholarship qualifying examinations such as the National Merit Scholarship Examinations in the United States.

Administering the examination :

After the examination has been carefully planned and constructed--the next important step is to arrange for the administration of the test. Often teachers do not regard this part of the operation as important as some of the other details but, in fact, if the test is to perform the task it is assigned to do--the environment and arrangements should be carefully thought out. Some simple reminders might be helpful.

1. The examination room should be well-lighted and well-ventilated.
2. Instructions of what the students' are supposed to do should be clear and precise.
3. The proctor should indicate how much time is to be allowed for completion of the test and, perhaps, how they should gauge their time.
4. Any unusual characteristics of the test should be explained.
5. The scoring formula should be explained, i.e., if the score is to be determined by the total number of correct answers - or some special formula will be used to discourage guessing on the part of students.
6. Students should be informed if they are to choose the one BEST answer--or if there are multiple answers in each item.
7. Outside distractions and interruptions should be minimized.
8. Seats should be arranged to minimize the possibility of cheating. If the class is large and the examining room is small, an alternate form of the examination should be made with both the items and the responses scrambled.

9. The proctors should circulate around the room keeping a watchful eye to make sure that the test is accurately measuring the work of each individual student. Proctors, obviously, should move slowly and unobtrusively to avoid distracting the students.

Scoring and analyzing the examination :

As soon as possible after the test has been administered it should be scored and returned to the students. Ordinarily it is a good idea to use the test, not only as a measuring instrument but also as a teaching device. If the answers are recorded on a separate answer sheet the tests should be returned to the students so they can analyze their mistakes (and help to point out weaknesses/ambiguities in the test itself) and be able to associate the questions with the correct answers.

Teachers who are using objective tests for the first time will probably be surprised at the range of scores especially if the test is truly discriminating. While there is no specific agreement on this by experts in test construction – it is probably fair to say that the MEDIAN score should be someplace between 50% and 60% of the total possible score. Until a test has been given a sufficient number of times to establish reasonable levels of performance – the teacher should NOT set any ABSOLUTE “breaks” for determining grades. The frequency distribution should be recorded on a sheet to note the distribution. Example :

Score	No. of students
20-	0
19-	1
18-	2
17-	0
16-	2
15-	3
14-	4
13-	4
12-	6
11-	8
10-	4
9-	1
8-	0
7-	6
6-	0
5-	2
4-	1
Total	44

The range of scores extends from 4 to 19 on a twenty-item test. The MEDIAN would be 12 which is 60% of the highest possible score.

After looking at the range of scores the teacher may want to translate these into grades. This process must necessarily be done at the discretion of the teacher depending upon the percentage of each grade the teacher thinks necessary - or allowances he may want to make if the test seemed particularly easy or difficult. After determining the percentage of each grade he wants to assign it is a good idea to look for "natural breaks" that may appear in the range of scores. In the above example it would appear that 6 and 17 would make good breaking points because

they distinguish more definitely between the students' performance than any other place on the scale. If it is possible (within the previously established percentages of grades) – these “natural breaks” should be used. If the teacher decides on a grade division of

10% A's

20% B's

50% C's

25% D's

5% F's

the breaks would fall at 18 and above = A (6.8 %)

14 thru 17 = B (20.4%)

8 thru 13 = C (52.3%)

7 = D (13.6%)

6 and below = F (6.8 %)

After grades have been assigned the teacher should look at the test to see what it indicated about student performance and the attainment of the objectives of the course. It is helpful if the teacher has the following information to work from: (1) the percentage of students who answered the question incorrectly (called the “item difficulty” or “index of difficulty”) and (2) an index which shows the ability of individual items to differentiate between individual students possessing much of some characteristic (skill, knowledge, etc.) from students who have little of the characteristic (this is called the index of discrimination and shows the discriminating power of an item). An easy way for the teacher to determine the index of discrimination is by dividing the test papers into three even groups: one-third with the highest scores – one-third in the middle range of scores and one-third in the lowest level of scores. By determining the proportion of students from the “high group” who answered a given item correctly and subtracting the low group proportion the result is the INDEX OF DISCRIMINATION. If the class is very small – it is permissible to divide the papers into two rather than three groups.

By rank order listing of the index of difficulty for each item, the test maker can see the range of difficulty of individual test items. This information is valuable because it enables the teacher to have a balance of easy and difficult items for future examinations. It also is valuable because it can help the teacher to decide whether to be lenient or harsh in assigning letter grades to students in a given test. The index of discrimination will help the teacher to evaluate whether an item "works" by seeing whether the best performing students choose the right response and the weaker students select a foil or distracter. Both of these indices are important for letting the teacher assess the effectiveness of the test. They also provide comparative data about items which might be used in future tests for comparative data.

It is also interesting and informative to note particular items for the popularity of each response. This is a further step in ITEM ANALYSIS which helps to make known some common misconceptions of students should influence the way the material is taught the next time through. It also reveals any foils or distracters which are never chosen by students. In future tests these ineffective foils can be replaced by ones that might be more tempting to the weaker students.

Two other factors should be mentioned in relation to test analysis : Test VALIDITY and Test RELIABILITY. There are a number of ways by which one can test the validity of a given test. One might be to see whether it tended to give a measure of student performance which is similar to the measure given by some other comparable evaluation instrument. Another would be to look at the test objectively to see whether the questions were really significant and whether they tested the skills suggested by the objectives. This check for validity might, again, have some influence on the teacher who must translate the scores into grades. It is possible, too, that this analysis would help to improve future examinations by pointing out certain weaknesses and flaws.

Test reliability refers to the degree to which a student would obtain the same score if a similar test or another form of the same test were given again to the student (assuming no additional learning etc.). Test reliability might also mean the consistency with which the test would separate students from one time to another. If the test did NOT show consistency in these instances—it could not be considered to be a good instrument for making judgments about a student's work. For the classroom teacher—probably the best indicator of reliability is the range of scores on a test. If the range is extensive—it may usually be assumed that the instrument is reliable. If a test is not reliable, its validity also diminishes.

GLOSSARY OF TERMS

(Taken from "A Glossary of Measurement Terms",
California Test Bureau, Monterey, California)

Achievement Test – A test that measures the level a pupil has achieved in one or more subject fields or in the general aspects of schooling.

Correction for guessing – A technique devised to adjust the number of right answers for a test by using some portion of the wrong answers. The assumption is made that if an examinee guesses on an objective test, the number of right and wrong answers resulting will be proportional to the number of alternate responses in each item. Common correction formulas are :

$$S = R - \frac{W}{n-1}$$

S = Corrected Scores

R = Number of Right answers

W = Number of Wrong answers

n = Number of Choices

Correlation coefficient – This is the most commonly used measure of relationship between paired facts or of the tendency of two or more variables, or attributes to go hand-in-hand. It ranges from -1.00 for perfect negative relationship through 0.00 for none or pure chance to +1.00 for perfect positive relationship.

Criterion – A standard, norm, or judgment used as a basis for quantitative or qualitative comparison.

Difficulty value – The per cent of a specific group who answer a test item correctly.

Discriminating power – The ability of a test item to differentiate between individuals possessing much of some characteristic (skill, knowledge, attitude) from those possessing little of that characteristic.

Frequency distribution – A tabulation of scores from high to low (or low to high) showing the number of persons who obtain each score or group of scores.

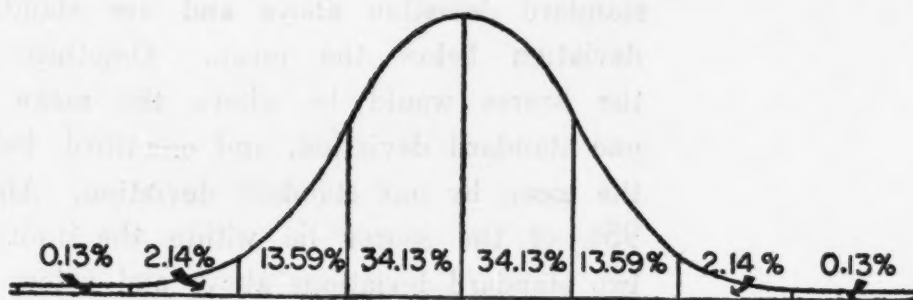
Item analysis – Any of several methods used in test construction to determine how well a given item discriminates among individuals differing in some characteristic. The effectiveness of a test item depends upon three factors: (1) the validity of the item in regards to curriculum content and educational objectives; (2) the discriminating power of the item in regards to validity and internal consistency; (3) the difficulty of the item (see Criterion).

Mean – The sum of a set of scores divided by the number of scores.

Median – The middle score in a set of ranked scores. It is the point above or below which an equal number of ranked scores lie, it corresponds to the 50th percentile.

Normal distribution curve – A derived curve based on the assumption that variations from the mean are by chance. It is bell-shaped in form and adopted as true because of its repeated recurrence in the frequency distributions of sets of measurements of human characteristics in psychology and education.

It has many useful mathematical properties. In a normal distribution curve, scores are distributed symmetrically about the mean, as many cases at various equal distances above the mean as below the mean, and with cases concentrated near the average and decreasing in frequency the further one departs from it.



Percentile – One of the 99 point scores that divide a ranked distribution into groups, each of which contains 1/100 of the scores. A percentile rank is a person's rank in a standard group of 100 persons representative of the full range of the normative population. If a person obtains a percentile rank of 70, his standing is regarded as equalling or surpassing 70 per cent of the normative group on which the group was standardized; a percentile rank score of 70 may also be interpreted to mean that 30 per cent of the normative group excel this person's test performance.

Range – The difference reflected by noting the lowest and the highest scores obtained on a test by some group.

Reliability – The degree to which a pupil would obtain the same score if the test were readministered to the pupil (assuming no additional learning, practice effects, etc.); **trustworthiness of scores.**

Standard deviation – (See Normal distribution curve). It is a statistic used to express the extent of the deviations from the mean for the distribution. It is obtained by taking the square root of the mean of the squares of the deviations from the mean of a distribution. If the group tested is a normal one their scores, if plotted graphically, would yield a normal distribution curve. Approximately two-thirds (68.3%) of the scores would lie within the limits of one standard deviation above and one standard deviation below the mean. One-third of the scores would be above the mean by one standard deviation, and one-third below the mean by one standard deviation. About 95% of the scores lie within the limits of two standard deviations above and below the mean. About 99.7% of the cases lie within the limits of three standard deviations above and below the mean.

Validity – A test is valid to the extent that we know what it measures or predicts or that it measures what it purports to measure.

ផែនការប្រតិបត្តិការបណ្តាញ ៩៩-៧១ របស់ក្រុមការងារ គណៈកម្មាធិការ គណៈកម្មាធិការ
បណ្តាញប្រតិបត្តិការ ផ្នែកបណ្តាញប្រតិបត្តិការ គណៈ

การเผยแพร่เอกสาร

เอกสารนี้เป็นเอกสารต้นฉบับในชุด "การศึกษาแห่งประเทศไทย" ซึ่งแจกจ่าย
ถึงหน่วยงานต่าง ๆ ที่เกี่ยวข้องในประเทศไทย และเผยแพร่ไปยังประเทศต่าง ๆ ทั่วโลก เพื่อ
ให้เป็นไปตามความมุ่งหมายในหน้าที่การงานของเอกสารนี้แล้ว ผู้ที่ประสงค์จะได้รับเอกสารนี้เพิ่ม
เติมนิ (ถ้ามีเอกสารเหลืออยู่) โปรดติดต่อตามสถานที่ดังต่อไปนี้โดยทันที

สำนักงานวางแผนการศึกษา	กระทรวงศึกษาธิการ ประเทศไทย
สำนักงานสภาการศึกษาแห่งชาติ	สำนักงานสหรัฐอเมริกา ประเทศไทย
Educational Planning Project,	MSU - USOM
Institute for International	APO San Francisco 96346
	Studies in Education,
	College of Education
	Michigan State University
	E. Lansing, Michigan 48823

DISTRIBUTION

This is the first title in the **Education in Thailand** series. A limited number of copies will be distributed to appropriate offices in Thailand and throughout the world in accordance with the objectives stated on the first pages of this booklet. Additional single copies (as long as the supply lasts) may be obtained by addressing requests to one of the following offices:

Educational Planning Office,	Ministry of Education
National Education Council	Bangkok, Thailand
Educational Planning Project	Prime Minister's Office
Institute for International	Bangkok, Thailand
	MSU-USOM
	APO San Francisco 96346
	Studies in Education,
	College of Education
	Michigan State University
	E. Lansing, Michigan 48823